1994-09

Evaluation of the Haworth-Newman avionics display readability scale

Chiappetti, Charles F.

Monterey, California. Naval Postgraduate School

AD-A286 127

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

EVALUATION OF THE HAWORTH-NEWMAN AVIONICS DISPLAY READABILITY SCALE

by

Charles F. Chiappetti

September 1994

Thesis Advisor: Judith H. Lind

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE		Form Approved OMB No 0704

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 1994	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE Evaluation of the Haworth-Newman Avionics Display Readability Scale (U)	5. FUNDING NUMBERS
6. AUTHOR(S) Chiappetti, Charles F	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution is unlimited	12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

This study investigates the suitability of the Haworth-Newman Display Readability Rating Scale as a performance-based test and evaluation tool. This evaluation has been necessary to determine if the scale actually measures display readability, and if consistent, reproducible results are attainable. Background information on the scale's development is presented along with a brief description of display readability characteristics. A technique for systematic degradation of display readability and a method of displaying degraded symbology sets is introduced. A flight simulation experiment was conducted to obtain performance data, Haworth-Newman readability ratings, and participants' written comments for each of the degraded symbology set levels. Five Naval test pilots attempted to maintain specified heading, altitude, and airspeed while utilizing the ten levels of symbology sets and then used the Haworth-Newman scale to rate the display readability for each. Experimental results are discussed and recommendations presented.

14. SUBJECT TERMS Avionics, Video Display Terminals, Legibility, HUD, Human Factors	15. NUMBER OF PAGES 71 16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited.

EVALUATION OF THE HAWORTH-NEWMAN AVIONICS DISPLAY READABILITY SCALE

by

Charles F. Chiappetti
Lieutenant, United States Navy
B.S., University of Kansas, 1987

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN AERONAUTICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
September 1994

Author: [redacted]
Charles F. Chiappetti

Approved by: [redacted]
Judith H. Lind, Thesis Advisor

[redacted]
E. R. Wood, Second Reader

[redacted]
Daniel J. Collins, Chairman
Department of Aeronautics and Astronautics

ABSTRACT

This study investigates the suitability of the Haworth-Newman Display Readability Rating Scale as a performance-based test and evaluation tool. This evaluation has been necessary to determine if the scale actually measures display readability, and if consistent, reproducible results are attainable. Background information on the scale's development is presented along with a brief description of display readability characteristics. A technique for systematic degradation of display readability and a method of displaying degraded symbology sets is introduced. A flight simulation experiment was conducted to obtain performance data, Haworth-Newman readability ratings, and participants' written comments for each of the degraded symbology set levels. Five Naval test pilots attempted to maintain specified heading, altitude, and airspeed while utilizing the ten levels of symbology sets and then used the Haworth-Newman scale to rate the display readability for each. Experimental results are discussed and recommendations presented.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

AOB	Angle of bank
FLSIM	Flight simulator
HMD	Helmet-mounted display
HUD	Head-up display
nm	Nautical mile
SGI	Silicon Graphics, Inc.
VAPS	Virtual Applications Prototyping System
VSI	Vertical speed indicator

I. INTRODUCTION

This study is rooted in the dynamic and ever-expanding area of avionics display symbology. In the present environment of decreasing budgets and increasing reliance on technologic innovation, the field of avionics has become a focal point for government and industrial investigation. The Boeing 777 with its "glass cockpit" and fly-by-wire design represents the latest in a long string of commercial designs that place considerable emphasis on avionics and displays. On the military side, recent budgetary and policy decisions have brought the F/A-18 D/E to the forefront of the United States Navy aircraft inventory. This multipurpose aircraft achieves its great flexibility in missions and roles through the extensive use of avionics and associated displays. These two examples point the way to the future.

The rapid growth and implementation of avionics systems have resulted in numerous unanswered questions relating to ergonomics, human factors, and man-machine interfaces. Of particular interest to this study is the area of display symbology comparisons, as these comparisons pertain to head-up displays (HUDs) and helmet-mounted displays (HMDs). A fundamental problem in this area has been the lack of an objective, performance-based evaluation criterion. A display readability rating scale, intended to serve as a performance-based evaluation tool, has been proposed to solve this problem (Haworth, 1993). The purpose of this study is to determine the suitability of that proposed scale, as a step toward its use in military test and evaluation programs.

A. DEVELOPMENT OF AVIATION DISPLAYS

Modern aviation displays can be traced back to the birth of military aviation. The placement of the first machine gun on World War I vintage aircraft led to sighting problems for early pilots. As technology developed the iron gunsights of these machine guns were replaced. By World War II the reflecting gunsight was the primary target designation device. This later evolved into a collimated display that allowed the pilot to focus on both the target and the sight, rather than having one appear blurred or doubled, resulting in the lead-compensating optical sight. Essential flight information was added to the display format to aid the pilot in maintaining an eyes-out orientation. As display technology matured increasingly more information has been added to the format resulting in the modern HUD. (Haworth, 1993, p. 1)

The information provided on a HUD is coded as symbols. These symbols can be letters and numbers (alphanumeric symbols) or can be geometric shapes and icons (graphical symbols). Generally the individual symbols are combined into a symbol set, designed to provide the necessary information rapidly and without confusion.

Development of head-up and head-down symbol sets is an ad hoc process. Each airframe has a unique set, with varying formats, contents, and symbols as required for its mission. Surveys of pilots familiar with the platform and mission usually serve as the basis of these designs. Today, considering budgetary restraints and the need for joint cooperative research and development of aircraft systems, this approach to display design is outdated.

B. HAWORTH-NEWMAN RATING SCALES

A hurdle to achieving efficient and standardized symbol sets and formats has been the lack of objective performance-based grading criteria with which symbology designs can be evaluated. Haworth and Newman have proposed two rating scales, the Display Readability Rating Scale and the Display Controllability Rating Scale, which could serve as these criteria. These two scales were developed to gather information on two fundamental flight display issues: "Can the pilot determine the value of a specific parameter, such as airspeed?; and can the display be used to control that variable?" (Haworth, 1993, p. 7). This study will focus solely on the readability issue and determination of the suitability of the Haworth-Newman Display Readability Rating Scale for test and evaluation purposes.

Based on the well-established Cooper-Harper Handling Qualities Rating Scale (Figure 1) used by test pilots for over 20 years, the Display Readability Rating Scale (Figure 2) utilizes a decision-tree process to guide the user through a series of questions. The answers lead the user to a set of three subalternatives which ultimately result in a numeric rating from 1-10. This choice of a decision tree and final ten user ratings stems from the early work of Cooper and Harper (Cooper, 1969, pp. 10, 15).

The early work of Cooper and Harper in devising a pilot rating scale to evaluate the handling qualities of aircraft led them to the use of four broad categories within which to describe these qualities. These categories are:

1. Satisfactory: no improvement required.
2. Unsatisfactory but tolerable: adequate for the task but improvement desirable.
3. Unacceptable: not suitable for the task but aircraft still controllable.
4. Uncontrollable: unsuitable for any task.

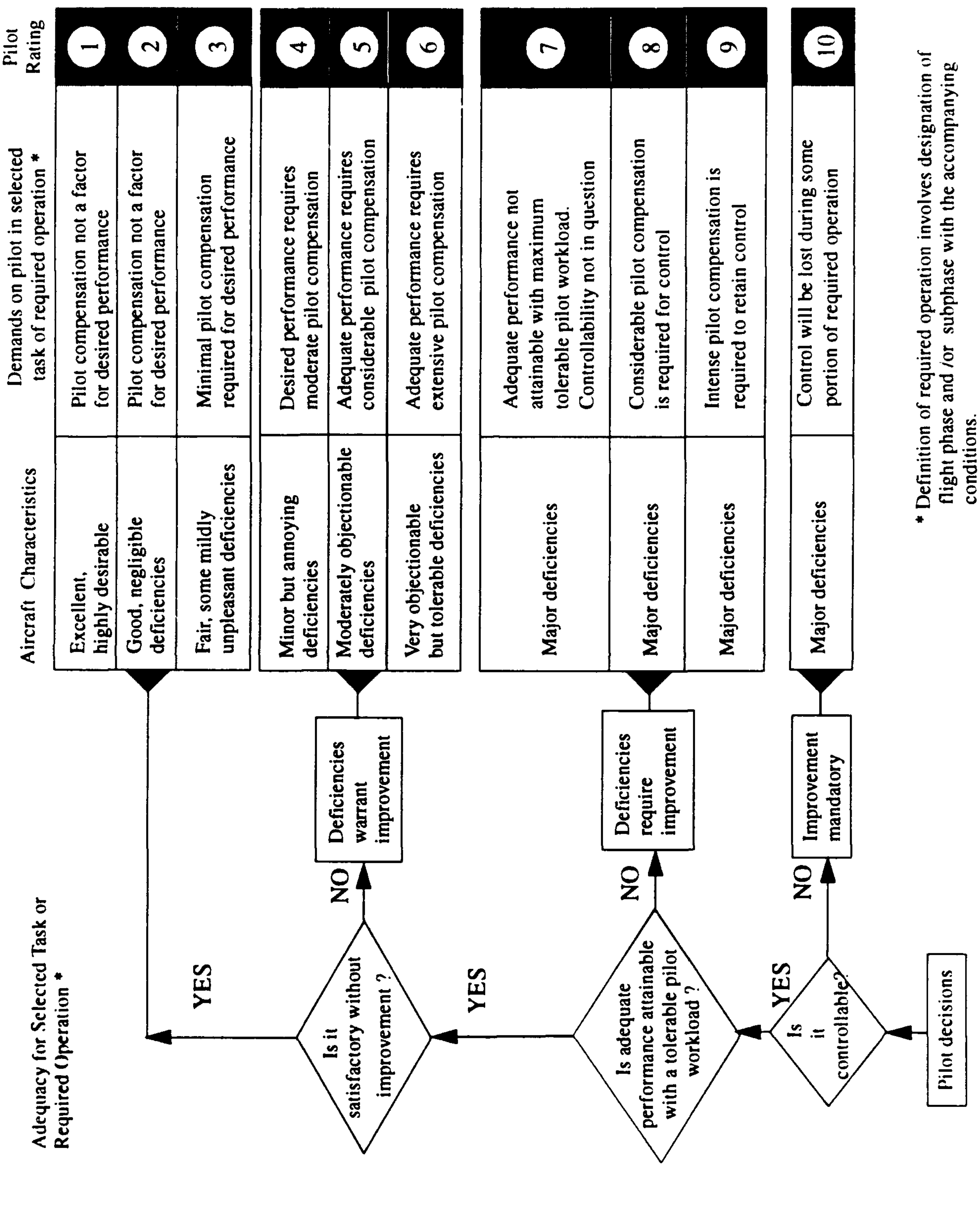

Figure 1. Cooper-Harper Handling Qualities Rating Scale (From Cooper 1969)

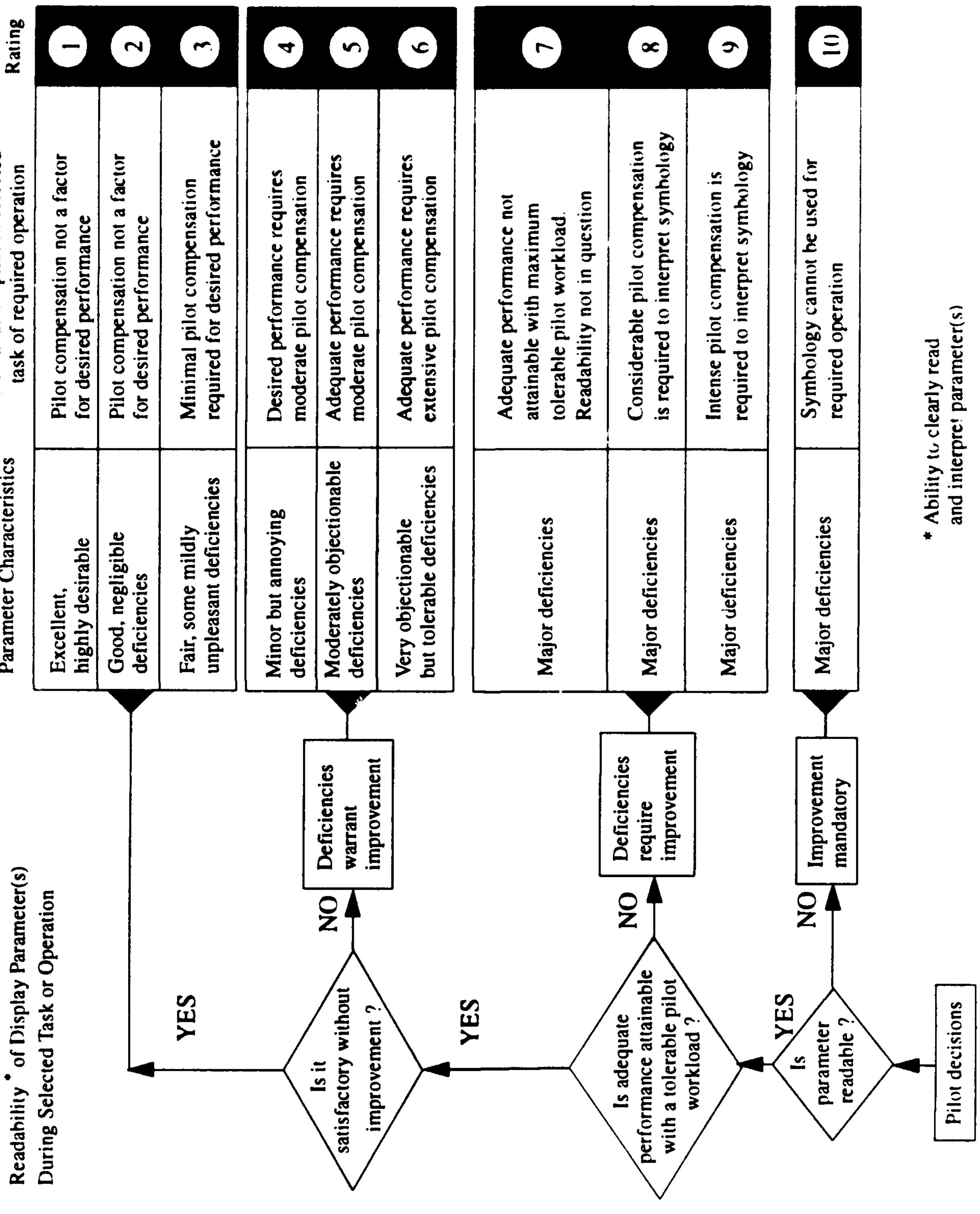

Figure 2. Haworth-Newman Display Readability Rating Scale (From Haworth 1993)

The following three questions help the pilot place the system into one of the four categories

1. Is the vehicle controllable?
2. Is adequate performance attainable?
3. Is system quality satisfactory without improvement?

These three questions form the basis for the Cooper-Harper scale decision tree. By separating the three upper categories into three subdivisions it was felt that an adequate spread would be achieved. Additional subdivision of the final category was not considered to be of value. These elements form the ten ratings available with the scale. The Display Readability Rating Scale adopts these same categories and Cooper-Harper decision tree process.

It is important that users of the scale understand and utilize the category definitions and make the decisions listed on the left of the scale. Inappropriate results will occur if only the numeric values and their descriptions are used. The important boundaries between 3-4, 6-7, and 9-10 cannot be distinguished from the descriptions alone.

Another important aspect of the scale is the emphasis placed on pilot performance. Two levels of performance, adequate and desired, must be defined by the experimenter. These two performance levels form the foundation of the rating system, as they will directly determine which numeric rating will be given.

Lastly, key definitions found in the decision tree must be considered by users, along with the numeric descriptions. For the Display Readability Rating Scale specifically, these are:

1. Readability.
2. Workload.
3. Pilot Compensation.

Readability is defined for the scale as "Ability to clearly read and interpret parameter(s)" (Haworth, 1993, p. 8). *Workload* and pilot performance were recognized to be interdependent concepts by Cooper and Harper. Thus performance could not be determined independent of workload considerations. The Cooper-Harper definition of workload "... is intended to convey the amount of effort and attention, both physical and mental, that the pilot must provide to attain a given level of performance" (Cooper, 1969, p. 12). *Pilot compensation* is a function of the increase in workload required to improve performance, considering task difficulty and required precision. Compensation can be thought of as the additional effort and attention required to maintain performance in the face of less favorable characteristics (Cooper, 1969, p. 13).

C. GOALS AND OBJECTIVES

The goal of this study has been to determine the suitability of the Haworth-Newman Display Readability Rating Scale as a test and evaluation tool, as suggested by Loran Haworth of the NASA-Ames Research Facility, Moffett Field, California. This evaluation has been necessary to determine if the scale actually measures readability, and if consistent, reproducible results are attainable through use of the scale. Haworth considered that a satisfactory result would be a standard deviation of 1 with respect to the expected rating value. However, with a limited sample size of study participants, an acceptable result would be if the ratings fall into the four broad categories of the scale.

A series of objectives were met during the completion of this study. These objectives are covered briefly here and described in detail in subsequent sections.

First, a method was required to display a set of symbols with systematically varied readability levels. Symbols and formats developed for an earlier Naval Postgraduate School Aeronautical Engineering thesis study were used. The apparatus consisted of two commercially-available software packages: an interactive graphic animation package, and a flight simulation program. These programs were run on a computer provided by the Naval Postgraduate School's Visualization Laboratory.

Second, a technique was needed to vary the symbols physically so that readability varied systematically on a ten-point scale. A simple dynamic HUD format was created using the graphics software and coupled with the flight simulation software. The HUD's heading, altitude, and airspeed readability were degraded over a ten-level scale by placing a mask of varying density over their respective readouts.

Finally, participants were gathered to evaluate the readability of the ten levels of HUD clarity. They were tasked to maintain 360° heading, 500 feet altitude, and 200 knots airspeed for 3 minutes in a simulated instrument flight profile. They performed this task once with each level of degraded HUD. After each run they rated the HUD's readability using the Display Readability Rating Scale. Both pilot performance data and subjective ratings were gathered. Data analysis, results, conclusions, and recommendations are presented in the remainder of this thesis.

D. SCOPE

The rapid advancement of avionics display technology has outpaced the test and evaluation communities' ability to compare different symbology designs and formats objectively. This study has explored the readability aspects of avionics displays by using

test pilots to evaluate a proposed objective performance-based rating scale. These pilots already possessed the knowledge needed to use performance-based scales and were experienced in the evaluation process. A readily-reproducible experiment, in which systematically-degraded readability levels of display formats were used, has been carried out.

Limitations of available experimental hardware did not permit addressing controllability issues, as these issues pertain to display systems. No attempt has been made to investigate the effect of symbol placement with respect to pilot field of view. Additionally, no attempt has been made to investigate display formats *per se* or their optimization.

II. DISPLAY CHARACTERISTICS

This study is concerned with the concept of display readability. There are two interrelated aspects to this concept. First, legibility is generally defined as a display characteristic that affects the ability to identify a single character or symbol. On the other hand, readability is a display characteristic that affects cognitive processes used to understand the meaning of symbols, such as when reading text (Spenkelink, 1993, p. 254).

The human visual system and its ability to process information have been studied intensively by the scientific community. A vast body of knowledge presently exists, but the rapid pace of electronics development continues to foster a vigorous research effort. Much of this current research deals with human vision as it relates to military displays and to display quality.

Human visual perception is rooted in phenomena in three domains: light, space, and time. Interactions of these three phenomena determine what the eye and brain perceive (Spenkelink, 1993, p. 250). Display quality is therefore a multidimensional concept. The complex interactions of these three phenomena preclude a single definition of display quality. The literature, in fact, contains numerous definitions of quality (Snyder, 1985 and Roufs, 1980).

Typically, display quality is measured in two ways: (1) physical measurements of the display characteristics, or (2) perceived quality based on human observation. Physical measurements of the display usually are made by engineers and pertain to advances in

display design or to other engineering aspects. Human observation approaches usually are taken by social scientists to determine how well the human can use a given display to perform a particular task.

Numerous factors in the three domains can affect display quality. Five such aspects relating to display quality will be briefly discussed.

Resolution refers to the smallest detail that can be shown on a visual display. Typically, *resolution* is expressed as the number of total lines which are available on a cathode ray tube for illumination or by the number of lines per unit distance (Cushman, 1991, p. 102). Shurtleff (1980, p. 65) demonstrated that a minimum of 10 lines per symbol height are required to achieve a high level (99%) of symbol identification accuracy. *Resolution* and *symbol size* are interrelated and, to maintain this 99% identification accuracy with respect to number of lines per symbol height, a minimum symbol size of 12 to 16 minutes of arc is required (Shurtleff, 1980, p. 65).

Brightness is generally considered to be the subjective sensation of various light levels emitted or reflected from an object. The related term for the physical measure of light is *luminance* which has units of foot-Lamberts (fL) (Bylander, 1979, p. 57). *Brightness* is a major determiner of the *contrast* between the display and its immediate surroundings and is responsible for the level of adaptation of the visual system (Spenkelink, 1993, p. 253). Displays having higher levels of *luminance* allow finer details to be seen on the display. Recommended *brightness* values for black and white cathode

ray tube displays are 10 to 50 fL. Recommended values for color displays in daytime are 20 to 90 fL, and for nighttime 2 to 9 fL (Lind, 1981, pp. 27 and 37).

Symbol size is primarily described by the symbol's subtended arc-angle and by the symbol width-to-height ratio. The arc-angle (α) is given by: $\tan(\alpha) = \frac{H}{D}$; H = symbol height; D = distance from the display to the eye in the same units as H (Bylander, 1979, p. 51). Shurtleff (1980, p. 41) states that a symbol width-to-height ratio of 75% is recommended for cathode ray tube displays.

Contrast is a measure of the difference in either *luminance* or *color* of an object of interest and the background on which it is displayed. *Luminance contrast* is defined by Cushman (1991, p. 96) to be the ratio of the luminance of an object (Lo) to its background (Lb). This ratio may be expressed as: $\frac{Lo}{Lb}$: 1 if $Lo > Lb$ or $\frac{Lb}{Lo}$: 1 if $Lb > Lo$. For example if Lo = 15 fL and Lb = 5 fL the contrast would be in a ratio of 3:1. Studies conducted by Howell (1959), Crock (1954), and Shurtleff (1979) indicate an increase in symbol identification accuracy with an increase in contrast ratio. *Color contrast* is the relationship between the symbol color and the background color.

A complex interaction exists between *contrast*, *symbol size*, and *luminance*. Shurtleff (1980, p. 33) reports that a *contrast* ratio as low as 2:1 may be used when *luminance* is greater than 10 fL and *symbol size* is greater than 10 minutes of arc. However if *luminance* is low (0.01 fL to 0.1 fL) the recommended *contrast* must be on the order of 5:1 with symbols greater than 20 arc-minutes and on the order of 18:1 if symbols are less than 20 arc-minutes. When color displays are considered, a *contrast* ratio between 20:1 and 30:1 is recommended (Lind, 1981, p. 37).

Sharpness describes the relationship between the edges of a symbol and the background. It can be thought of as how clearly the symbol edge is distinguished from its background. Physical attributes of the cathode ray tube which affect *sharpness* are the *resolution*, pixel size, pixel shape, and inter-pixel spacing. As manufacturing technology continues to decrease pixel size and spacing, displayed symbol edges appear more distinct and smoother to the eye. *Contrast* is also a factor in how sharp a symbol appears. Increased *contrast* increases the edge distinction between symbols and the background.

To achieve the desired ten readability levels of this study, symbol *contrast* and *sharpness* were systematically degraded. This approach was implemented by placing a software generated mask over the displayed symbols, as is described in Chapter III.

III. METHODOLOGY

The Haworth-Newman scale is a decision tree matrix leading to a ten-point scale consisting of levels of display acceptability. These levels range from (1) satisfactory performance, highly desirable, to (10) unreadable, major deficiency. Evaluating the scale thus requires developing display formats that vary in readability systematically on a ten-point scale. A HUD symbology set was chosen as the basic test element. This set was altered by overlaying it with a mask which varied in density from (1) no mask to (10) total obscuration of the symbols. This resulted in a linear spectrum of readability, to cover the Haworth-Newman scale. That is, as discussed in Chapter II, contrast and sharpness of the symbols that made up the format were systematically degraded from excellent (1) to unreadable (10). Aviators then evaluated the ten display levels using the Haworth-Newman scale, and their performance while using the various readability levels was monitored. Thus comparisons could be made between:

1. Known readability levels as determined by mask density.
2. Participants' judgments of readability using the Haworth-Newman scale.
3. Participants' measured performance levels while flying with each of the 10 readability levels of the symbol set.

A. EQUIPMENT

1. Hardware

The evaluation was conducted on a Silicon Graphics, Inc., 380/VGX graphics workstation. The machine includes eight 33 megahertz IP7 processors, each with 256

megabytes of random access memory. Peripheral equipment included a serial mouse used to simulate an aircraft stick, a keyboard used to simulate the throttle via successive depressions of the "t" key, and a 19-inch diagonal color monitor for the HUD symbols and the out-the-window scene.

2. Simulation Software

The basic HUD symbology set was designed using the Virtual Prototypes, Inc., Virtual Applications Prototyping System (VAPS). This software package allows for rapid graphical design implementation. It possesses a graphical user interface which eliminates the need for extensive computer graphics programming skills. An extensive set of linking tools allow this program to interface with many hardware components and C-based software packages.

A second program, the Virtual Prototypes, Inc., Flight Simulator (FLSIM), was used as the simulation platform. The HUD symbology set was linked to FLSIM and used as the primary flight instrumentation. FLSIM incorporates an out-the-window scene generation capability with reconfigurable aircraft flight dynamics for fixed-wing simulations. Because it is also designed with a graphical user interface it is fairly simple to reconfigure most aircraft parameters by point-and-click operations. Numerous modifications are permitted, including those to airframe parameters (e.g., center-of-gravity position, wingspan, wing area, weight, fuel load, control surface deflections, etc.), aircraft performance parameters (e.g., lift and drag curves, engine thrust schedule afterburner response, etc.), atmospheric conditions, and initial conditions (Marshall, 1993, p. 51).

B. BASIC SYMBOL AND FORMAT DESIGNS

The basic HUD symbology set used in this evaluation was designed by Marshall (Marshall, 1993). The set was originally used in experiments conducted to investigate wide-field-of-view HMD symbology (see Figure 3), and was designed to provide a simple functional set of fundamental flight data indicators.

Marshall's symbology set was modified to meet specific requirements of this study, as shown in Figure 4. The HUD format as used included:

1. An airspeed indicator with digital readout in the left half of the field of view.
2. An altitude indicator with digital readout and vertical speed indicator in the right half of the field of view.
3. A magnetic heading display and digital angle-of-bank indicator located above the center point of the display.

The HUD symbology and format are purposely simple and uncluttered. Criteria for satisfactory readability (as discussed in Chapter II) generally were met. Individual symbols incorporated in the display design comply with the general requirements of MIL-STD-1295A (MIL-STD-1295A, 1990). The design was also influenced by recommendations from the Naval Air Warfare Center Aircraft Division, Warminster, PA. No effort was made to optimize individual designs or overall layout. (Marshall, 1993, p. 52).

Marshall's experimental results suggest that a lateral separation angle between the airspeed and altimeter groups of between 40° and 60° produces the best pilot performance in this simulation environment. Thus a lateral separation angle of 50° is used throughout this evaluation.

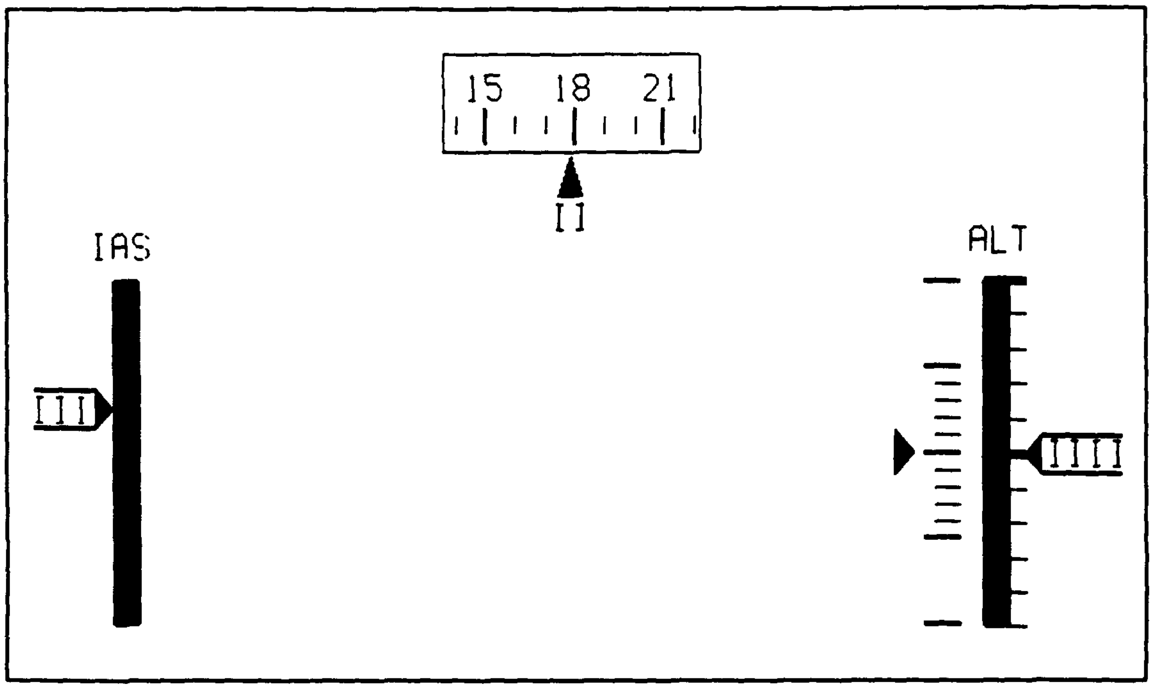

Figure 3. Wide–Field–of–View Symbology Set (From Marshall, 1993)

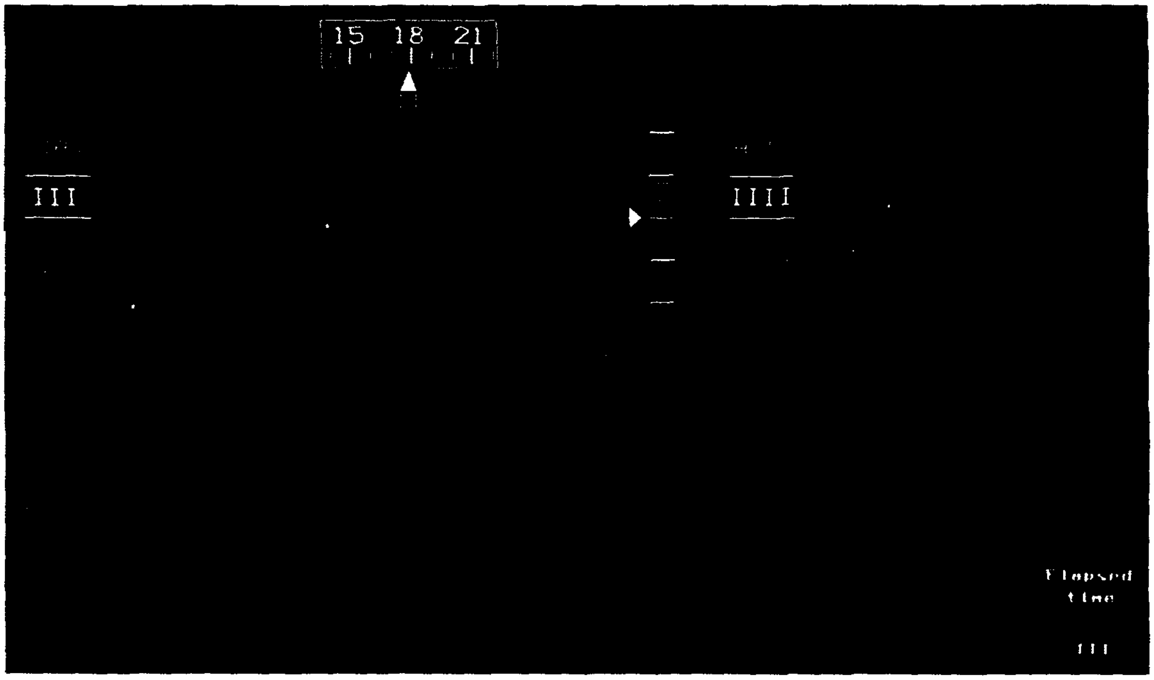

Figure 4. Modified Symbology Set

C. EXPERIMENTAL DESIGN

The Haworth-Newman scale, though a readability scale, is fundamentally linked to the task that is being performed. In order to validate this scale, a suitable task must be defined. Maintaining a basic instrument flight profile was chosen as the task. The question of display readability must also be addressed. The basic symbology set of heading, altitude, and airspeed formats were objectively and systematically degraded in a linear fashion, as described below. This degradation formed the basis of the readability evaluation.

The independent variable for the evaluations was the objective readability level of the heading, altitude, and airspeed displays, assumed to be a function of the degree to which the symbols were degraded by the mask, from (1) unmasked to (10) completely masked (unreadable). All other conditions remained the same. Each participant flew all ten evaluation flights and used all levels of symbol masking. Subjective readability ratings via the Haworth-Newman scale were obtained from each participant. Pilot performance was measured and compared to the subjective ratings. The dependent variables used to measure the pilot performance were deviations from the specified heading, altitude, and airspeed.

Each participant evaluated the ten levels of HUD readability, spanning the Haworth-Newman readability scale spectrum from 1 (excellent, highly desirable) to 10 (symbology cannot be used for required operation). Presentation of the ten displays was

randomized. Table 1 shows the order in which masking levels (readability levels) were presented to the participants.

Table 1: ORDER OF READABILITY PRESENTATION

	Order of Presentation									
	1	2	3	4	5	6	7	8	9	10
Participants	Masking Level Presented									
JO	2	8	5	4	1	9	3	7	10	6
JH	4	8	2	10	3	6	9	1	7	5
EE	4	9	1	6	3	5	2	7	10	8
DH	3	8	5	9	1	10	4	6	2	7
SG	2	7	5	9	4	8	3	10	1	6

1. Task and Simulator Parameters

The experimental design utilized for this evaluation is based on a study investigating HUD variations on basic flight performance conducted by Ercoline (Ercoline, 1990). Participants were tasked to fly a basic instrument profile, i.e., to maintain heading 360°, 500 feet, and 200 knots , for 180 seconds. This allows the aircraft to transit 10 nautical miles during the 180-second flight at the specified 200 knots.

The aircraft was perturbed from balanced flight over the desired flight path by means of wind vectors. These wind vectors are accessed in the FLSIM program via the atmospheric menu. A maximum of ten positional vectors can be defined at one time. User-defined values can be entered for north-south, east-west, and vertical velocity fc

each of the ten X-Y positions. Each position can further be subdivided in the vertical plane. User-defined velocities can be entered for sea level and up to five subsequent altitudes per position. This allows for 60 distinct wind vectors to provide the desired perturbation. Ercoline provided for perturbation by driving his altitude simulation with the sum of five sinusoids with different frequencies, amplitudes, and phases. The version of FLSIM utilized did not allow for input via data file; therefore wind variation was used to provide the desired motion.

Wind vectors were placed at 2.5, 4, 5.5, and 7 nautical miles ahead of the aircraft origination point. The line of wind vectors coincided with the desired flight path along heading 360°. This setup forced the aircraft off the target conditions and provided the sole component of pilot workload. Appendix A provides the wind settings utilized for this evaluation.

The wind simulations achieved the desired balance between attainable performance and aircraft perturbation. The 1.5 nm spacing provided approximately 27 seconds to allow the pilot to recognize and correct the perturbation. The single axes of perturbation and relatively small amplitudes did not require extreme control inputs for correction. These qualities were deemed desirable by the initial participants and subsequent evaluations showed that participants could achieve the desired performance goals, when the HUD format could be read.

2. Display Readability Degradation

VAPS was utilized to develop the desired ten-point symbol readability levels, using a modified version of Marshall's symbology set (Marshall, 1993) with alphanumerics changed from black to white (red value of 255, green value of 255, blue value of 255). This color change aided in producing the desired levels of degradation. The heading, altitude, and airspeed font was changed to vpi_font, a 13 x 23 pixel raster font provided with VAPS. These changes left a simple white, boldface display format suitable for contrast and sharpness degradation.

Symbol degradation was achieved by utilizing the texture function of VAPS. This function consists of a 16 x 16 pixel palette. Each pixel is mouse selectable to be on or off and assumes the currently selected color when applied in the workspace. This texture was applied as a mask over the numbers and symbols representing heading, altitude, and airspeed. The altitude and airspeed masks were approximately 3/8 x 3/4 inches and the heading mask was 3/8 x 2 1/2 inches as measured on the face of the monitor (see Figure 5 and Appendix B). The masks partly or completely obscured the symbols, resulting in various levels of symbol visibility on the HUD.

The mask color was yellow (red = 255, green = 250, blue = 0). This yellow-over-white color scheme provided a nearly uniform degradation over the spectrum of colors used by FLSIM as sky and terrain features. The underlying white numerics were judged to be slightly more visible through the mask when the displays were viewed on the dark green ground versus the blue of the sky.

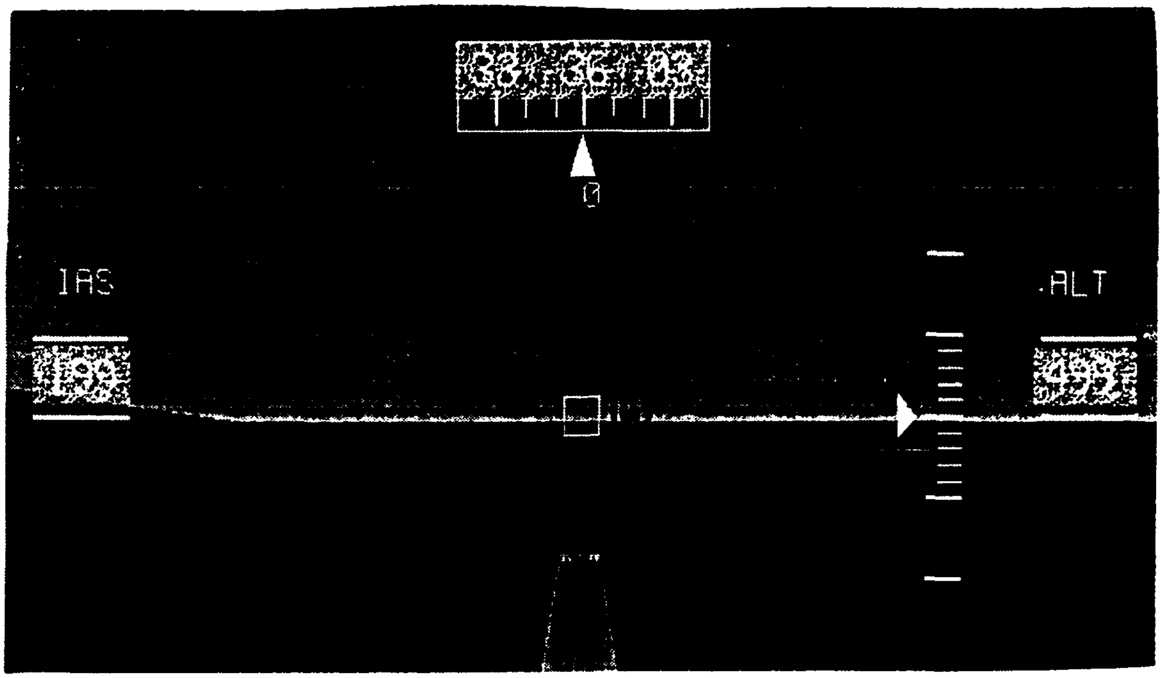

Figure 5. Example of Symbology Mask

Symbol degradation was achieved by systematically increasing by a linear amount the number of pixels turned on in the mask. Each step in the scale represents a 10% degradation. A mask level of 1 represents 100% symbol visibility or 0% degradation. A mask level of 2 represents 90% visibility or 10% degradation and so on. The number of mask pixels to be turned on was determined by subtracting the product of the total number of pixels and visibility percent from the total number of pixels: $256 - 256 \cdot x$, where x = visibility percent.

The 16 x 16 texture grid was subdivided into quadrants and the mask values randomly distributed within. For example, for rating 2 each quadrant received 6 random pixels and 2 quadrants received an extra pixel for 26 total pixels. The next successive mask level was built upon the previous level's design (e.g., for rating 3 the 51 pixels were

not randomly redistributed but instead 25 additional pixels were distributed onto the previous 26 pixels of design 2). Table 2 shows the values used; all values were rounded to the nearest whole number.

Table 2: READABILITY VS. MASK PIXEL NUMBER

Masking Level	Visibility Percentage	Mask Pixels (256 - 256·x)
1	100	0
2	90	26
3	80	51
4	70	77
5	60	102
6	50	128
7	40	154
8	30	179
9	20	205
10	10	230

D. SCENARIOS

All military aircraft evolutions have common mission segments, e.g., preflight, taxi, departure, navigation to mission area, mission phase, navigation from mission area, etc. Each mission segment has unique performance requirements. The task specified for this evaluation is similar to a low-level navigation flight profile.

Initial pilot evaluations formed the basis of the task-specific performance criteria used in this study. Performance was divided into two categories, adequate and desired. Adequate performance was defined to be maintaining ±10° heading, ±10 feet altitude, and

±10 knots, with respect to prescribed values: 360° heading, 500 feet altitude, and 200 knots airspeed. Desired performance was defined to be maintaining ±5° heading, ±5 feet, and ±5 knots. Similar methodology has been used elsewhere to collect and categorize performance data (Lind, 1980).

The simulation was conducted under daylight, visual meteorological conditions. Prevailing wind conditions have previously been described. The aircraft was capable of simulating speeds from 60 to 400 knots. The earth surface was essentially flat andfeatureless (see Figure 6). No depth or altitude cues were provided by the out-the-window scene, requiring participants to rely solely on their displayed instruments. The simulation was rendered in 24-bit color.

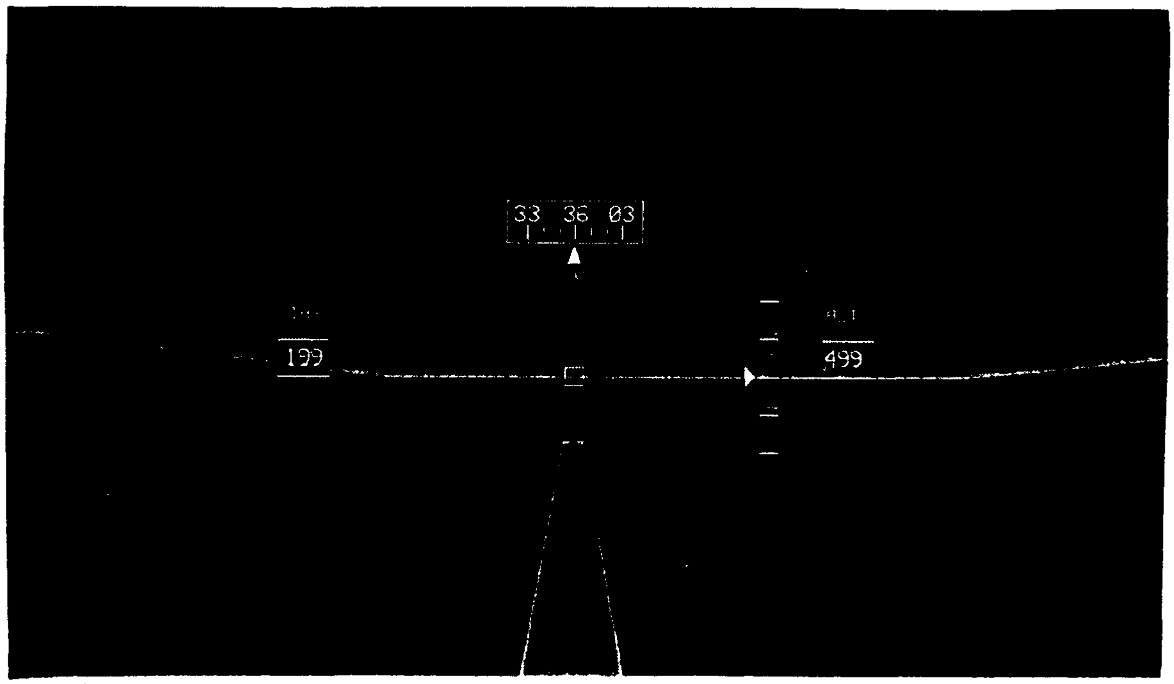

Figure 6. Displayed Out-the-Window Scene

E. EXPERIMENTAL CONDITIONS

All evaluations were conducted in the Naval Postgraduate School Visualization Laboratory. Participants were seated in front of the monitor in swiveling chairs that were adjustable for height. The keyboard and mouse were positioned for individual comfort. One bank of overhead fluorescent lights was illuminated. Screen glare was judged to be minimal and the additional lighting aided in keyboard utilization.

Each evaluation was observed by the experimenter, who was seated behind and to the left of the participant. Notes on the heading, altitude, and airspeed were taken on each run to help during the debriefing process. The experimenter called time checks at 1, 2, 2 1/2, and 3 minutes for each run. No verbal instructions were given as to altitude or airspeed corrections.

F. STUDY PARTICIPANTS

The Cooper-Harper and Haworth-Newman scale qualities discussed in Chapter I were paramount considerations when selecting participants for this investigation. Haworth and Newman raise the issue of whether operational pilots or test pilots should be used for system evaluations. Operational pilots have recent mission experience and their experience levels cover the complete spectrum from recent pilot graduates to seasoned veterans. A problem with their use is that they tend to have a predisposition to their particular aircraft's displays. These pilots also must be thoroughly trained in the use of the scale and in how to fly with non-standard displays. (Haworth, 1993, p. 11)

Test pilots are already familiar with the use of Cooper-Harper rating scales and have knowledge of the important definitions and descriptors used in the scales. They are experienced pilots and usually have broad exposure to various platforms and displays. They are experienced with communicating to designers and engineers and can provide insight into any display or control problems (Haworth, 1993 p. 11). The limited time available for participant training and the completion of this study dictated the use of test pilots.

Five male pilots participated in this study. Each was a fully qualified Naval aviator. In addition, all were graduates of the Navy's Test Pilot School and had completed at least one tour of duty in the capacity of a test pilot. Four participants were currently students in the Naval Postgraduate School Aeronautical Engineering Department. The remaining participant was an instructor at the Navy's Aviation Safety School which is a resident program of instruction at the Naval Postgraduate School.

G. PROCEDURE

Participants were tested individually. Each participant completed a preflight questionnaire (Appendix C) to provide general background and personal information. Overall experience levels were ascertained as well as test pilot histories and individual HUD experience. Participants were then briefed on the upcoming sequence of events and the purpose of the study. The outline used for briefing purposes is included as Appendix D.

The Haworth-Newman scale was briefed in detail. The definition of readability (as specified on the scale in the lower right corner) was covered. Each node of the decision tree was explained, along with its accompanying pilot rating descriptions. Examples of display readability variation (see Figure 7) were shown on the computer monitor. The importance of the participants' written comments and thought processes was emphasized. The participants were then briefed on their task. Adequate and desired performance criteria were discussed. They were told that ten evaluations would be conducted with time in between to provide written remarks.

The simulation was then initialized and the participants were briefed on the controls and HUD display. The use of the mouse for pitch and roll input was discussed, along with the use of the letter "t" for throttle inputs. The simulation had a slight discontinuity when it was initially released from static to dynamic state: the throttle would sometimes drop to approximately 0 %. This discrepancy was demonstrated and the participants were allowed to experience this during their practice flights. The HUD layout was reviewed and the function and limits of each item discussed.

The participants were then allowed to practice flying the simulator. Initially they familiarized themselves with the overall layout and sensitivity of the controls. They then practiced constant altitude, constant airspeed flight. Next throttle changes were introduced, followed by return to a constant altitude and airspeed condition. Finally, 3-minute practice runs were conducted. When the participant was able to maintain consistently adequate performance the practice was complete and data runs commenced.

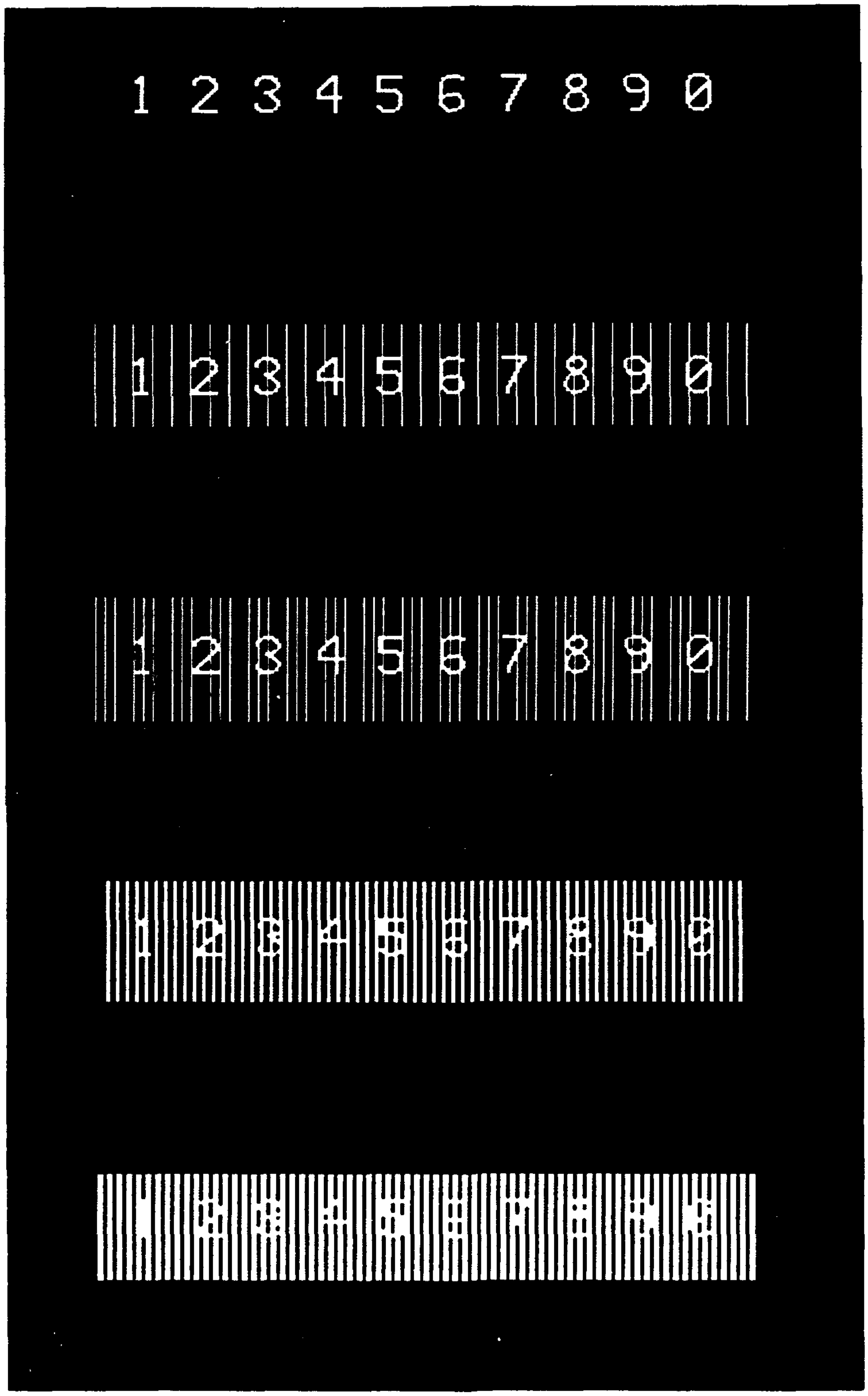

Figure 7. Examples of Display Readability

Prior to each data run the simulation was initialized to 360° heading, 500 feet, and 200 knots. The appropriate HUD format masking level (Table 1) was selected by the experimenter by means of a keyboard selection. The participant then positioned the keyboard, mouse, and monitor for individual comfort. The simulation was released and the participant attempted to maintain the desired performance criteria. The experimenter called out time checks at 1, 2, 2 1/2, and 3 minutes. The simulation was frozen at 3 minutes. This procedure was repeated ten times with each participant. All participants evaluated the same ten HUD symbol masking levels, presented in random order. Aircraft heading, altitude, and airspeed were sampled at 1 Hz and stored in a data file for later retrieval and analysis.

Upon completion of a data run, the participant evaluated the observed level of HUD readability using the Haworth-Newman scale and assigned an overall rating from the ten-point scale. Each participant was allowed as much time as desired to complete written comments.

IV. DATA COLLECTION, ANALYSIS, AND RESULTS

A. PARTICIPANT SUBJECTIVE RESPONSE DATA

1. Data Collection

At the end of each masking level evaluation the participant was given a copy of the Haworth-Newman Display Readability Rating Scale and asked to evaluate the display and provide a rating. Written remarks were also gathered at this time. Table 3 shows the Haworth-Newman readability ratings provided by the participants for each of the mask levels evaluated.

Table 3: SUBJECTIVE READABILITY RATINGS

	Masking Level Evaluated									
	1	2	3	4	5	6	7	8	9	10
Participant	Haworth-Newman Readability Rating Assigned									
JO	3	3	1	3	5	5	6	9	8	10
JH	4	5	6	4	3	6	7	9	8	10
EE	2	4	5	4	5	10	10	10	10	10
DH	2	5	5	2	6	7	9	10	9	10
SG	2	3	2	3	6	3	8	9	8	10
Mean	2.6	4	3.8	3.2	5	6.2	8	9.4	8.6	10
Variance	0.6	0.8	3.7	0.5	1.2	5.3	2	0.2	0.6	0
Std. Dev.	0.8	0.9	1.9	0.7	1.1	2.3	1.4	0.5	0.8	0

2. Data Analysis

The arithmetic mean, variance, and standard deviation of the assigned ratings was calculated for each of the masking levels. These results are at the bottom of Table 3. A plot of the expected values for the ten masking levels is provided in Figure 8, along with the means and variance of the assigned ratings. Dashed lines on either side of the expected values represent ±1 rating level around those values.

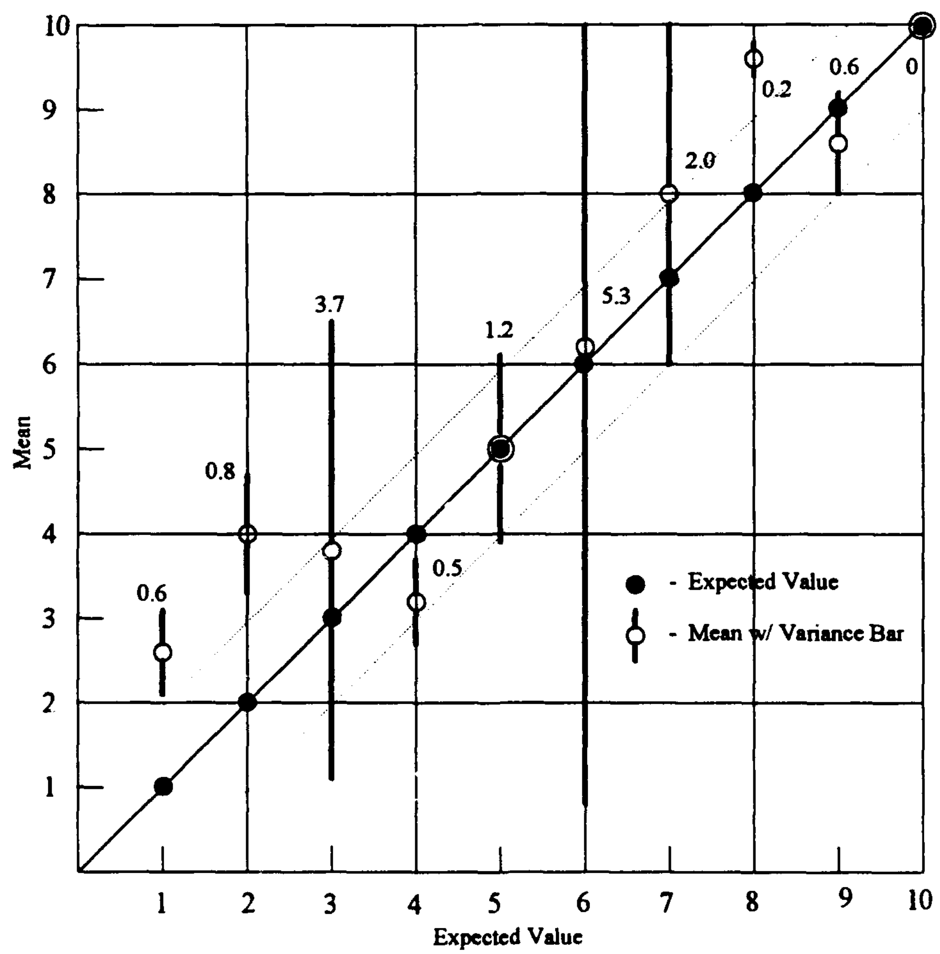

Figure 8. Expected Values Versus Means of Assigned Values for Readability Levels

3. Quantitative Results

The data presented in Table 3 and Figure 8 show the numeric results of this study. There is a strong correlation between the expected values (mask level) and the participants' assigned readability rating values, especially in the lower two-thirds of the scale (ratings 4 through 10). Seven out of ten means fell within ± 1 rating level of the expected value. Mask level 10 showed the strongest correlation, with all participants assigning a rating of 10.

The rating group consisting of mask levels 7 through 9 (representing the "Deficiencies require improvement" section of the scale) showed inconsistent results, with mask level 8 receiving a less favorable readability rating than 9. This is attributed to the masks' pixel distribution which produced strong curving features that tended to degrade severely numbers with curved shapes (2, 3, 5, 6, 8, 9, 0).

Mask levels 4 through 6 ("Deficiencies warrant improvement") arguably had the strongest correlation of the three major rating groups. The assigned ratings were the closest to the expected values. The exceptions are in level 6 where participant EE assigned a rating of 10 and SG assigned a 3. However, EE assigned a 10 for each mask level from 6 through 10. He determined that the legibility of these masks was so degraded that they were unsuitable for controlling the required parameters of the simulation, and he thus assigned a 10 rating to all of them. This assignment was not based on the readability of the display. His comments reflect that the symbols were readable with increasing levels of concentration and were generally consistent with ratings 6 through 10.

Participant SG assigned a rating of 3 to the mask level of 6. His comments reflect the decreased readability of the symbols, but he found that this made him concentrate more on the displays. This increase in attention was deemed desirable and thus a higher rating was assigned.

Mask levels 1 through 3 ("Excellent," "Good," "Fair") showed the least strong correlation in terms of the mean versus the expected values. But this group had the third and fourth smallest variations and standard deviations (level 1 and 2 respectively). Furthermore, the participants consistently rated this group the most readable. That is, the lowest rating (most readable) given by a participant appears in this group and the three ratings as a group reflect lower ratings. The exception is participant JH who assigned his lowest rating (3) to mask level 5.

4. Participants Comments

The participants' written comments are of greater importance than the numerical ratings, as they reveal the underlying causes of the assigned rating. For instance, the only rating of 1 was assigned by JO and this was for a mask level of 3. He commented that the small amount of yellow mask actually enhanced the contrast of the white numerals against both the dark green ground and the blue shades of the sky, and this was judged to be a desirable attribute. All the participants indicated a similar approval of a small amount of yellow masking. This is reflected in the comparatively high ratings assigned to mask level 4.

All participants reported that the white symbols were hard to read when they coincided with the pale blue-grey colorband which depicted the horizon. This condition occurred when the aircraft was in a straight and level attitude.

Participants stated that pilot workload was increased as the masking level increased. This is reflected in comments about concentration levels required to interpret symbology, and about how long attention was focused on a particular symbol and the subsequent breakdown of instrument scan. At masking levels of 7 through 9, participants forced the aircraft into a nose down attitude to place the masked symbology onto the dark green ground (which perceptibly increased the readability of the white numbers). This also allowed for interpretation of numbers based on the airspeed and altitude changes which occurred, that is, could they differentiate a number 3 from an 8 if the 3 changed to a 4 due to the forced change.

Participants reported that at higher masking levels (7 through 9) they could detect changes in the digital readout of the off-axis parameters with their peripheral vision but could not evaluate the change or the trend of the change. At the lower masking levels the trend could generally be identified with peripheral vision. An overall lack of aircraft trend information was indicated. At higher mask levels the participants would force a change in aircraft parameters to gain this trend information and at lower mask levels would have to remember previous values and then mentally determine trends. This caused an increase in pilot workload, but is a reflection of the HUD's informational content rather than symbol readability.

Participants experienced a noticeable "learning curve." Between three and six evaluations were required to master the use of the simulation interfaces and to anticipate the wind conditions that were experienced. Negative comments were made regarding the simulator dynamics and interfaces. The imposed limitation to ± 5% in throttle changes and lack of precise attitude control with the mouse were judged detrimental to the evaluations. The participants had a difficult time separating the less- than-ideal simulation handling qualities from their perceived ability to achieve adequate or desired performance.

The inability to provide real-time performance feedback to the participants was a problem. Performance data from each evaluation was stored in a data file but was not available for participant use. Access to this data would have helped separate simulator hardware inadequacies from actual participant performance.

Finally, the definition of readability as used in the scale received comment. It was felt that the word "clearly" could lead to misleading ratings. For instance, a mask level of 6 could not be read clearly, but was judged to be readable enough to maintain performance requirements. A strict application of the definition would require a rating of 10.

B. PARTICIPANT PERFORMANCE DATA

1. Data Collection

Performance data from each participant's masking level evaluations were stored in a computer-generated data file which recorded time, altitude, and airspeed approximately once per second. A total of 50 data files were generated, each with approximately 180 observations for each of the three measured parameters. The resulting

data files were reformatted to facilitate analysis using The Mathworks, Inc., Matlab computational software.

2. Data Analysis

The small number of participants limited the use of standard statistical analysis techniques. General performance trends were obtained by averaging each participant's airspeed and altitude data and then calculating the magnitude of the difference between those averages and the prescribed performance criteria (200 knots and 500 feet). These airspeed and altitude difference magnitudes (deviations) are presented in Table 4 (A/S Dev. and Alt. Dev., respectively). These data are graphically represented in Figures 9 and 10.

Table 4: AIRSPEED AND ALTITUDE DEVIATIONS

		Masking Level									
		1	2	3	4	5	6	7	8	9	10
JO	A/S Dev.	2.7	2	3.1	2.2	2.5	3.7	5.4	0.8	9.5	13.6
	Alt. Dev.	0.4	1.2	1.9	5.3	0.4	1.2	0	13.4	5.1	81.8
JH	A/S Dev.	2.1	4.1	19.8	179.1	8.3	3.9	0.7	10.4	5.9	17.2
	Alt. Dev.	7.3	4.2	3.8	67.4	4.9	2.4	0.8	131.1	4.3	68.1
EE	A/S Dev.	3.7	2.1	4.8	3.2	2.8	16	14.2	21.8	0.7	20.9
	Alt. Dev.	1.2	4.7	0.1	2.2	4.8	1.6	4.6	1.5	91	81.7
DH	A/S Dev.	2.3	5.9	0.9	3.9	0.6	3.2	4.1	21.6	5.1	21.9
	Alt. Dev.	3.9	26.9	9.4	4.8	28.8	13.9	14.9	29.9	15.9	31.8
SG	A/S Dev.	6.2	8.1	7.6	1.5	11.7	4.6	8.5	5.4	20.2	16.7
	Alt. Dev.	8.4	7.4	16.8	2.2	18.6	11.5	30.5	22.1	23.8	371.8
Mean	A/S Dev.	3.4	4.4	7.2	38	5.2	6.3	7.8	12	8.3	18.1
	Alt. Dev.	4.2	8.9	6.4	16.4	11.5	6.1	10.2	39.6	28.1	127.1

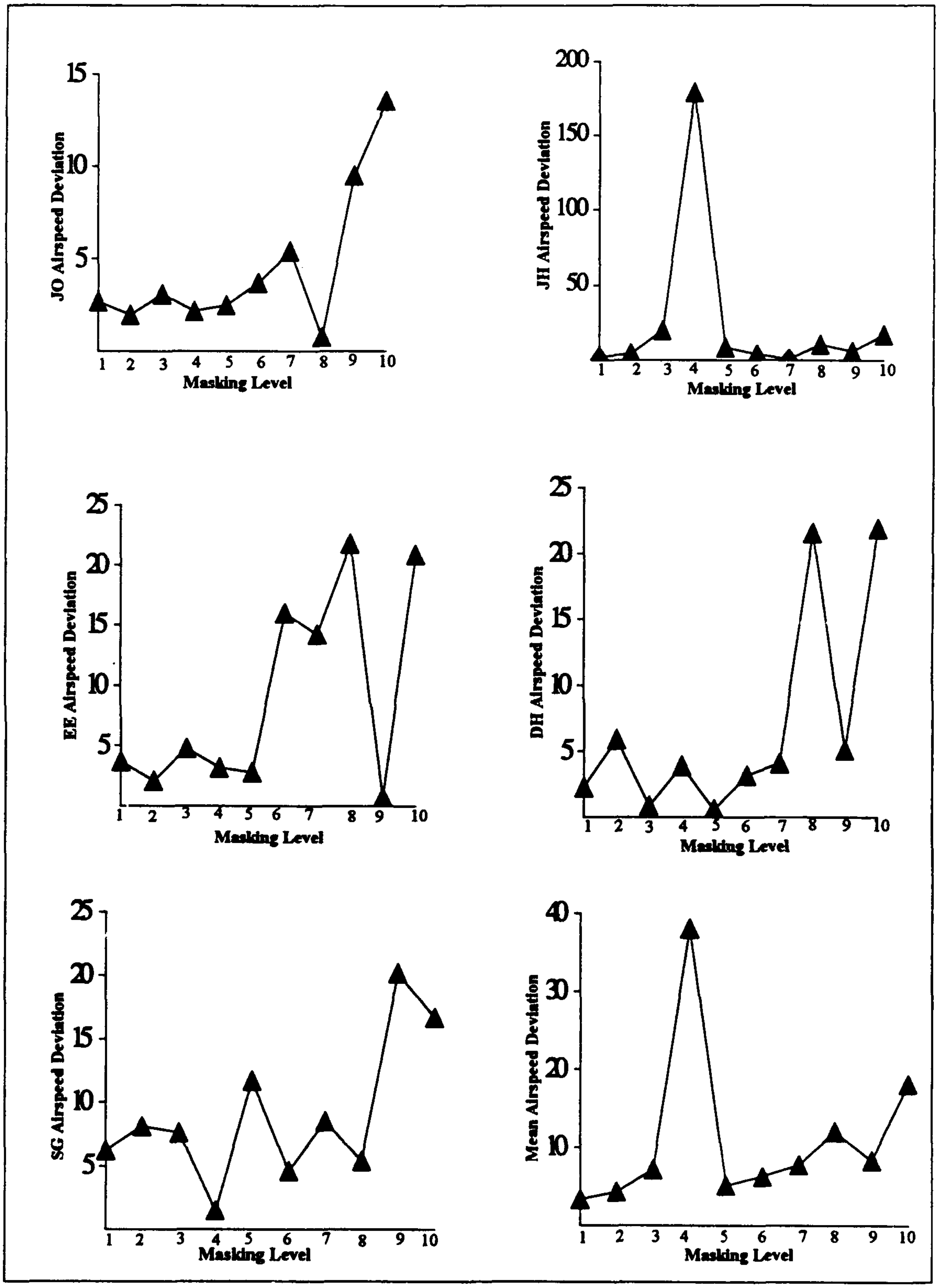

Figure 9. Airspeed Deviations versus Masking Levels

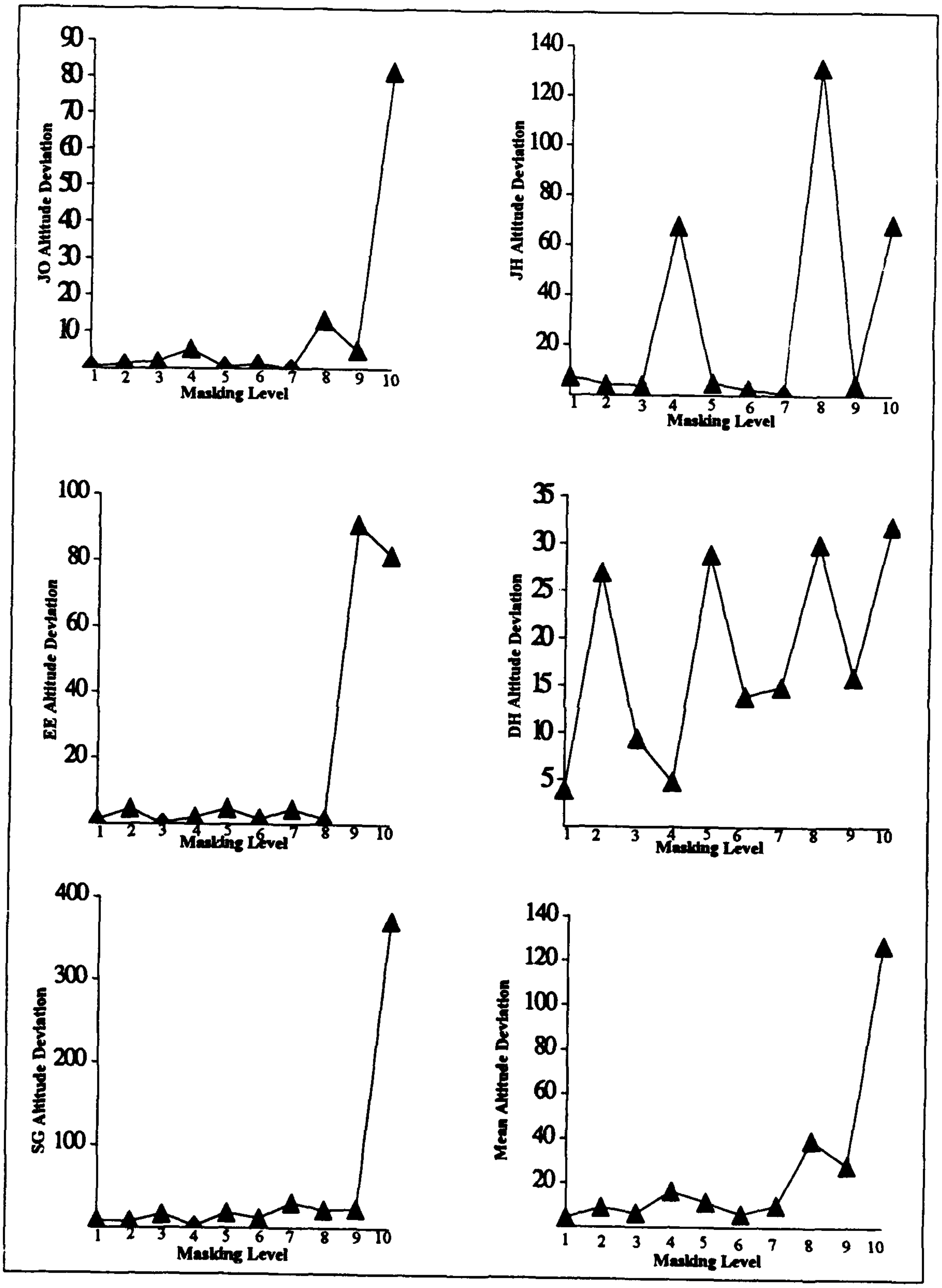

Figure 10. Altitude Deviations versus Masking Levels

3. Results

The data presented in Figures 9 and 10 show individual pilot performance deviations from the prescribed performance values of 200 knots and 500 feet. The trends for both sets are towards reduced pilot performance as the masking level increases or, conversely, as display readability decreases.

One anomaly in the data may be observed on Figure 9, for mean airspeed deviation. The peak for masking level 4 is due entirely to the performance of one participant, JH. He observed this level of masking on his first trial, and had considerable difficulty maintaining the required airspeed. Following that trial, his airspeed results were not significantly different from those of the other participants.

V. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

The goal of this study has been to determine the suitability of the Haworth-Newman Display Readability Rating Scale as a test and evaluation tool. It was therefore necessary to develop a method for displaying symbology sets and a technique which systematically varied the readability of those sets. A flight simulation experiment was conducted in which systematically degraded symbology sets were incorporated into a HUD format. Five Naval test pilots flew simulated missions using the ten levels of degraded symbols. They then used the Haworth-Newman scale to rate display readability. Based on the background research done for this study and on participants' performance, assigned ratings, and written remarks, three conclusions can be made.

First, as discussed in Chapter I, an objective, performance-based evaluation technique is needed to determine the readability levels of proposed aircraft displays. The Haworth-Newman Display Readability Rating Scale has been proposed to meet this need. Format and wording of this scale are consistent with the well-established Cooper-Harper Handling Qualities Rating Scale.

Second, the study reported in this thesis provides a preliminary indication that the Haworth-Newman scale may be a reliable measure of display readability. Although results are not conclusive due to the small number of participants included in the study,

performance trends and assigned ratings do provide sufficient evidence of the scale's value, as reported in Chapter IV. The scale appears to be flexible and possibly could be used to investigate specific readability issues (e.g., color contrast for individual symbols) or broader issues (such as the layout of entire display formats). Users obviously must receive adequate, standardized training on scale use and its key definitions. Their written comments are critical and must be considered in conjunction with the assigned numerical ratings.

Third, although the overall concept and implementation of the Haworth-Newman scale was well received by study participants, their comments (included in Chapter IV) indicate that the *definition* of readability used on the scale may be too restrictive: "Ability to clearly read and interpret parameters." Participants noted that the word "clearly" was to vague and could result in misleading ratings. Scale developers might consider including a more precise definition on the scale.

B. RECOMMENDATIONS

Several recommendations can be made. based on the study reported here. First, as noted above, the developers of the Haworth-Newman scale might consider a more precise definition of "readability" to minimize confusion for those using the scale.

Second, this study has been very limited. With only five participants, obtaining statistical significance was out of the question. Although the trends observed were in the right direction to indicate that the scale is applicable for test and evaluation, a full-blown validation program is recommended, using far more trained participants.

Third, any follow-on validation program should be conducted using more realistic experimental equipment. Simulation software should provide a more realistic out-the-window scene and simulate various luminance levels and visibility conditions. Simulation dynamics should be of high fidelity and input devices should be more representative of actual aircraft controls. Researchers should have the ability to give real-time performance feedback to the participants.

Fourth, the technique used to develop the ten levels of symbol readability for this study was based on systematic reduction of symbol contrast and sharpness by use of an obscuring mask. This technique was selected because it was relatively easy to implement on the equipment that was available. However, as discussed in Chapter IV, the colored mask resulted in varying levels of readability simply as a function of the kind of background (sky or terrain) against which symbols were viewed. Further studies should consider systematic variation of other parameters discussed in Chapter II to obtain precise levels of readability. *Display resolution, symbol luminance*, or *symbol size* might be considered candidates for such linear symbol degradation.

The Haworth-Newman Display Readability rating Scale shows great promise as a standardized test instrument for display design, to complement the Cooper-Harper scale for aircraft handling qualities. Thus, it is strongly urged that work continue on determination of this new scale's suitability for its intended purpose.

APPENDIX A: WIND COMPONENTS

The following table shows the wind components utilized in the evaluation. Positions are in nautical miles and are located along the 000° flight path.

Wind Components (kts)

Position (nm)	2.5	4	5.5	7
1000 ft.	20 up	15 hw	20 dw	15 tw
800 ft.	20 up	15 hw	20 dw	15 tw
600 ft.	20 up	15 hw	20 dw	15 tw
400 ft.	20 up	15 hw	20 dw	15 tw
200 ft.	20 up	15 hw	20 dw	15 tw
Sea level	20 up	15 hw	20 dw	15 tw

hw - head wind

tw - tail wind

up - up draft

dw - down draft

APPENDIX B: MASKING LEVELS

Figures 9 through 18 depict the ten masking levels used in this study. Each figure is a digitally reproduced image of the computer monitor with the FLSIM out-the-window scene and degraded HUD present. The original 19-inch diagonal monitor image was cropped to show the details of the degraded HUDs. The cropped images presented are close to true size.

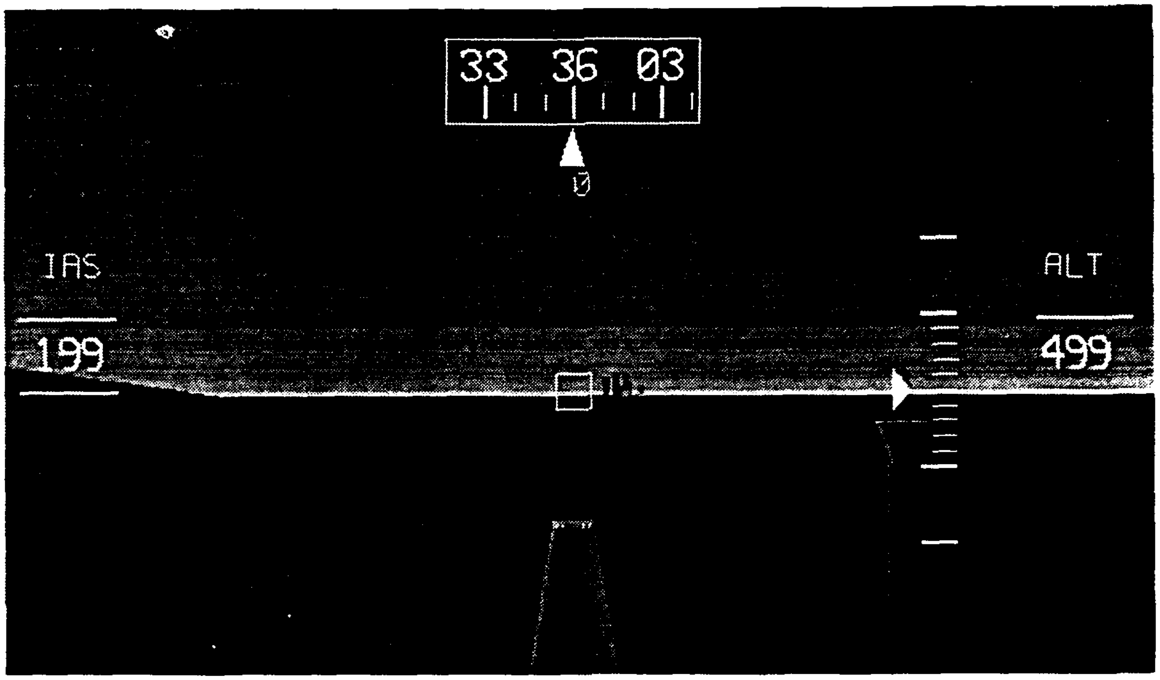

Figure 11. Mask Level 1

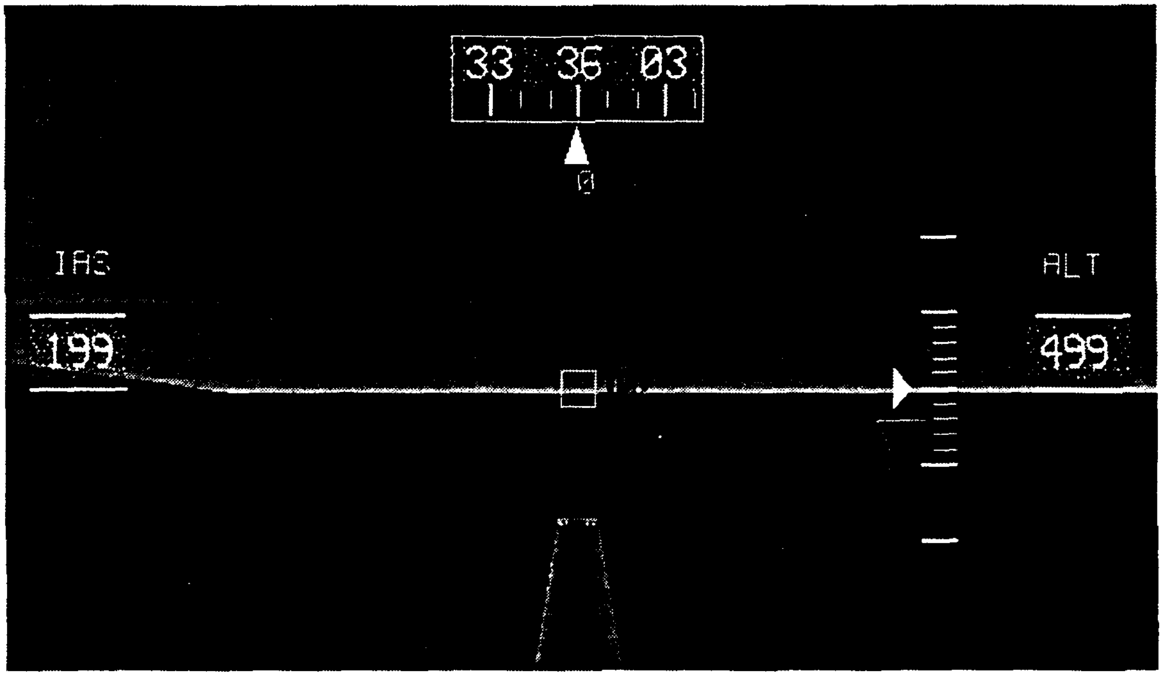

Figure 12. Mask Level 2

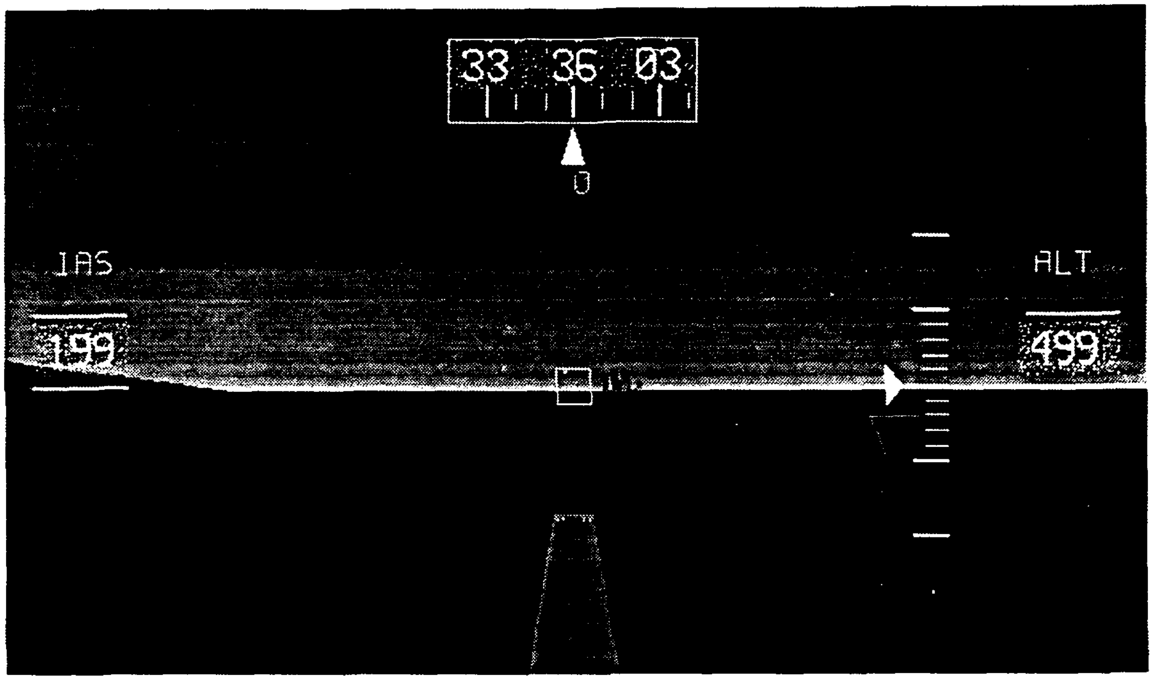

Figure 13. Mask Level 3

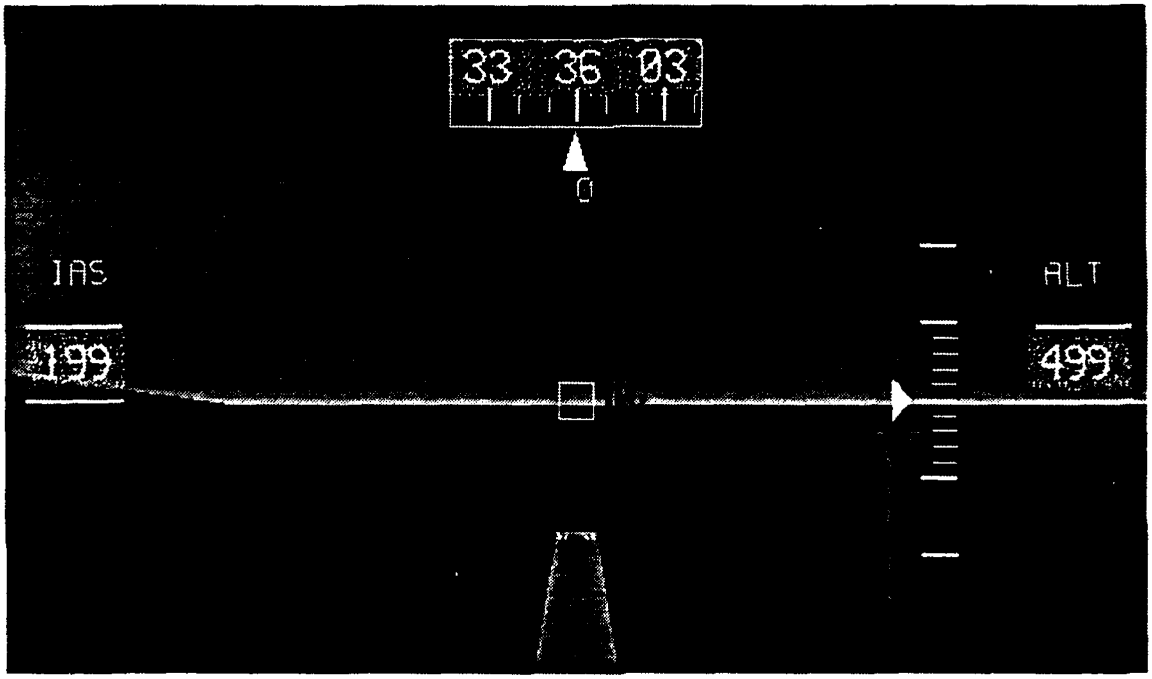

Figure 14. Mask Level 4

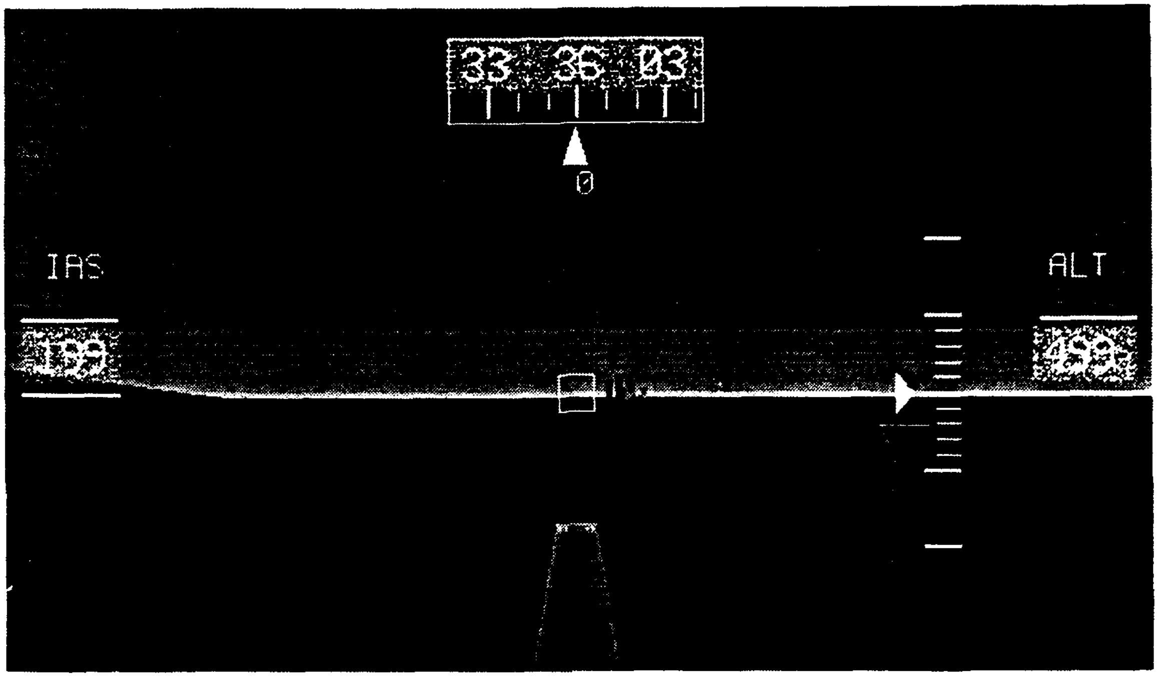

Figure 15. Mask Level 5

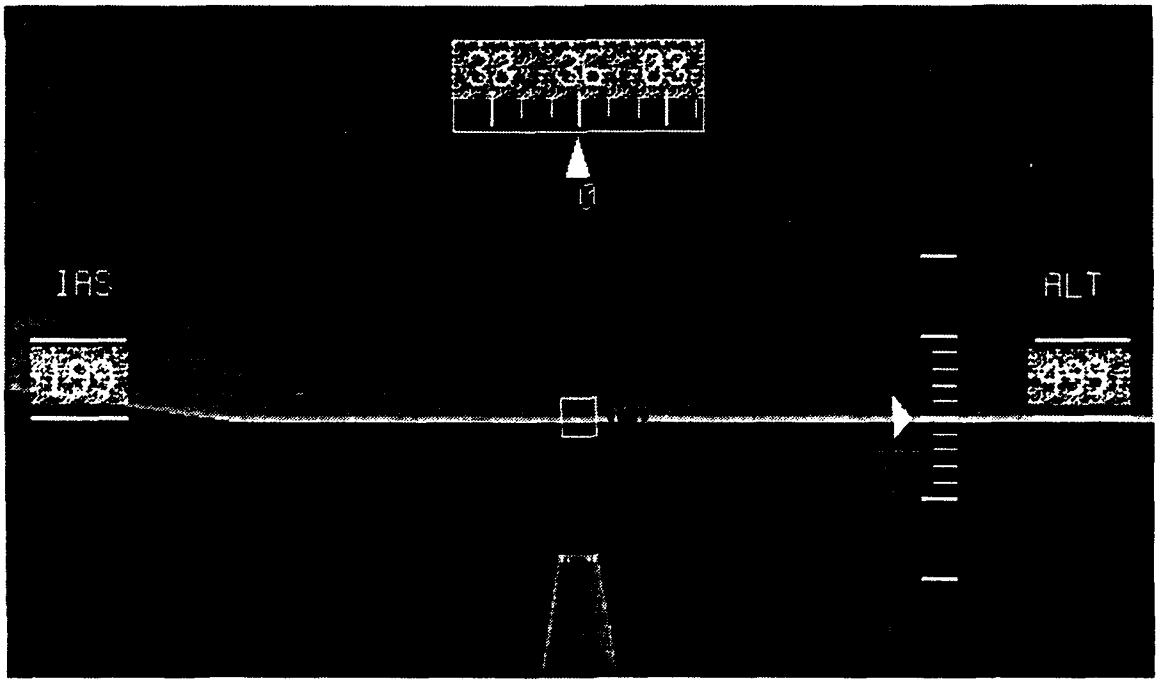

Figure 16. Mask Level 6

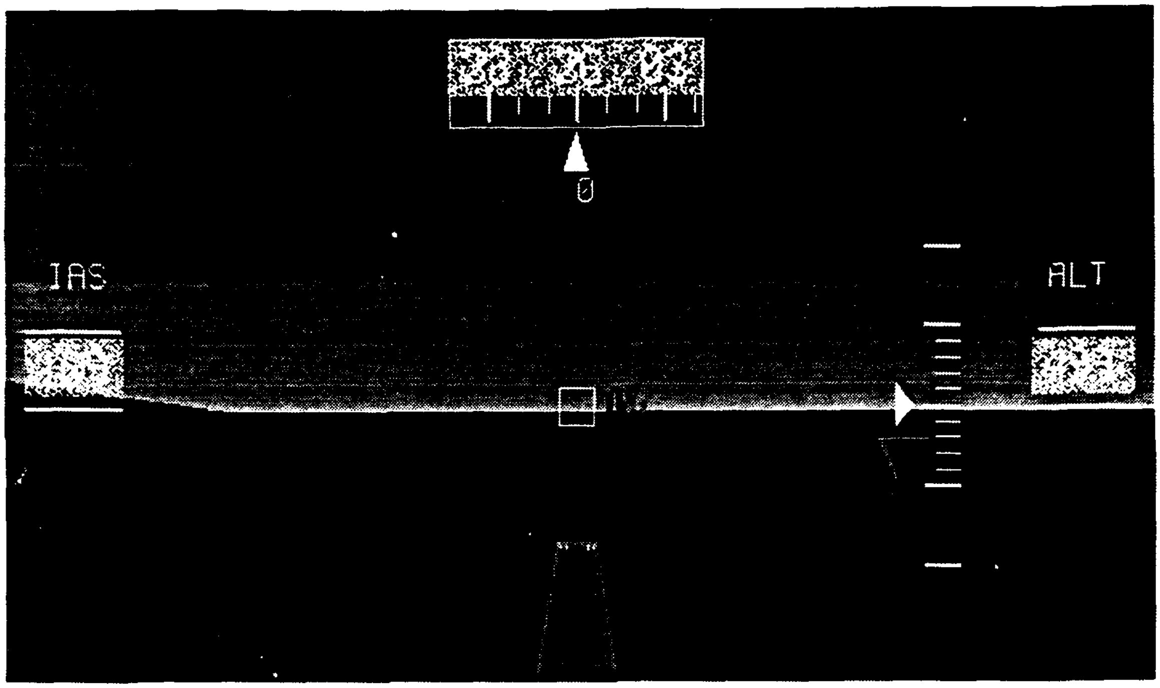

Figure 17. Mask Level 7

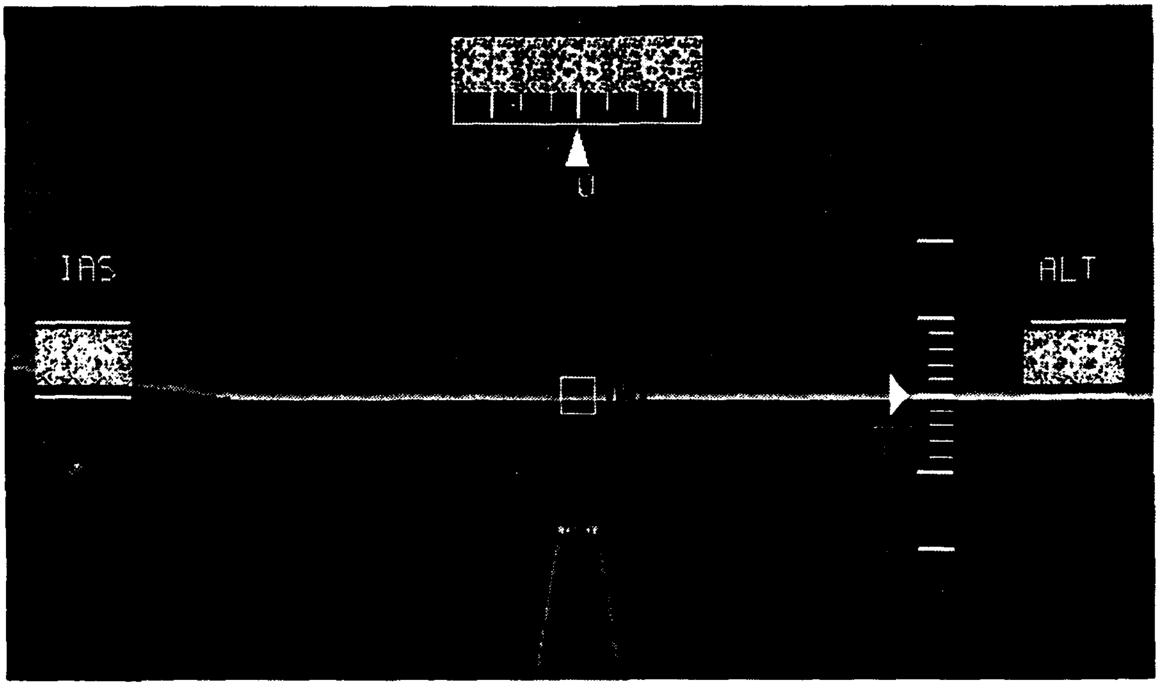

Figure 18. Mask Level 8

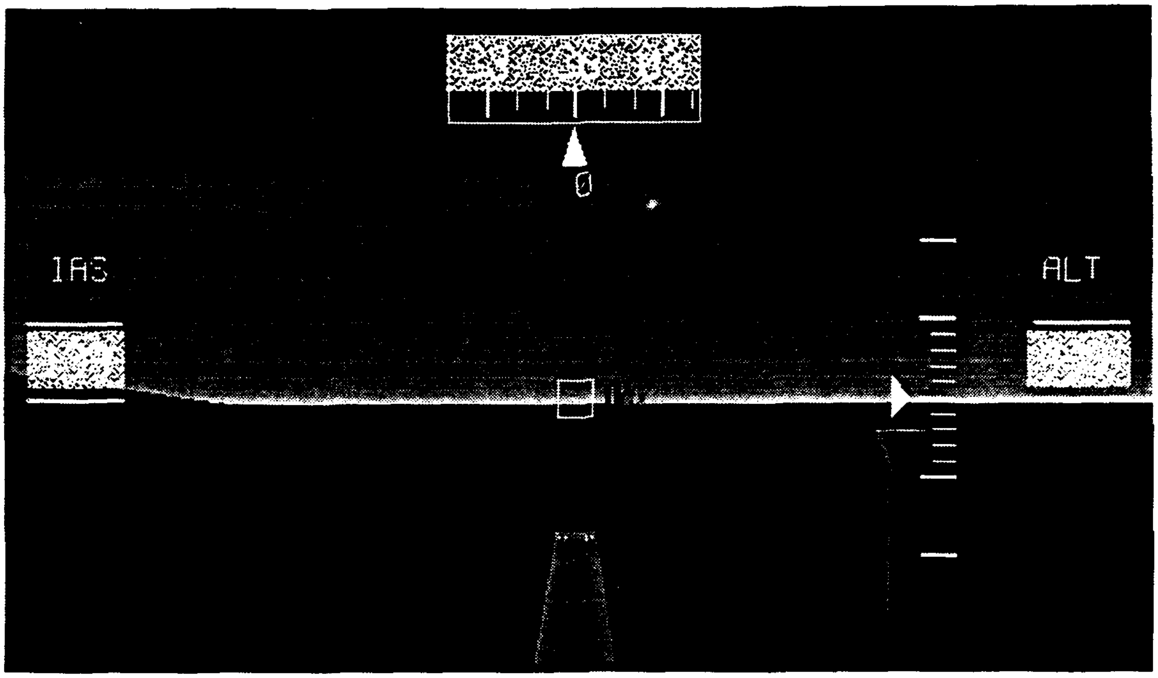

Figure 19. Mask Level 9

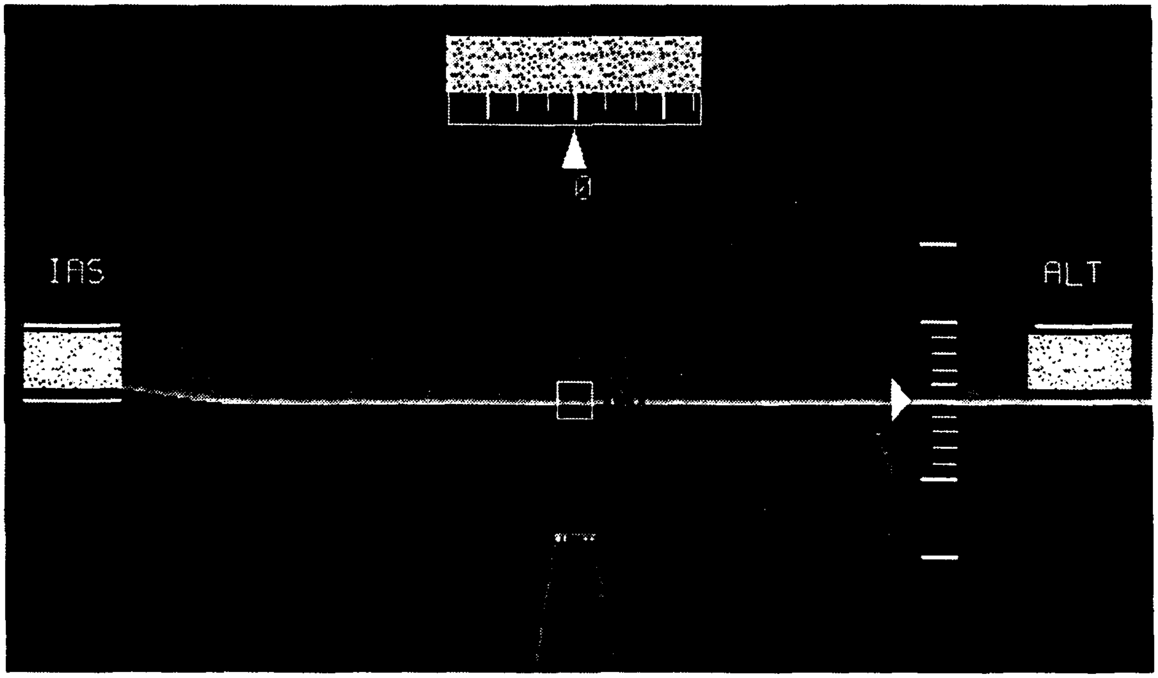

Figure 20. Mask Level 10

APPENDIX C: PARTICIPANT QUESTIONNAIRE

Rank/Name (first, last) ______________________________ Age ____ Sex ___

Service _________ Time in Service _______ yrs ______ mos

Designated Community: Rotary Wing / Fixed Wing (circle one)

Current Aircraft Type ________ Total Flight Hours _______

Months Since Last Flight _______

Flight Hour Summary (descending order, nearest 10 hours)

Aircraft Type _____ _____ _____ _____ _____ _____ _____

Hours _____ _____ _____ _____ _____ _____ _____

Qualified Test Pilot? Y/N TPS Grad Date _______ Last Test Flight ______

HUD Experience? Y/N if yes: Aircraft Type ______ HUD Flt Hrs ______

TO BE FILLED OUT BY RESEARCHER

Date / Time of Test 94 - _______ - ______ / _______

Visual Acuity 20 / _____ Eye Dominance R / L / N Handedness R / L / N

APPENDIX D: PARTICIPANT BRIEF

Sequence of Events

- Fill out questionnaire
- Conduct brief / answer questions
- Simulation training period ("fam flight")
- Conduct ten HUD evaluation runs
- Total time: approximately 1.5 hours

Purpose

- Validation of the Haworth-Newman display readability scale
- Scale is intended to be a real world tool in the evaluation of HUDs / HMDs

Haworth-Newman Scale Description

- Decision tree / ten point scale based on the Cooper-Harper flying quality scale
- Note upper left corner: scale is used to judge readability during selected task/operation
- Note lower right corner: readability is defined to be "Ability to clearly read and interpret parameter(s)"
- Show readability examples on computer
- Discuss decision tree logic and the ten rating descriptions
- Pilots' written remarks are critical components of the scale; why a particular value is assigned

Pilot Tasks

- Required to maintain 200 kts, 500 ft, 360° hdg for 180 seconds
- Adequate performance: ± 10 kts, ± 10 ft, ± 10°
- Desired performance: ± 5 kts, ± 5 ft, ± 5°
- Evaluate the HUD using the Haworth-Newman scale and provide written remarks
- Ten consecutive evaluations will be conducted with short breaks in between
- Pilot numerical ratings will be compared to pilot performance by use of data file of heading, airspeed, altitude stored at 1 Hz rate

Symbology Function and Location

- Heading tape and AOB readout
- Airspeed readout
- Altitude readout and VSI indicator
- "White box" at center of screen

Explain Control Inputs

- Throttle: increase "t," decrease "T," each change corresponds to ± 5%, drop to approximately 0% at beginning of simulation
- Pitch and roll: mouse

Familiarization Training

- Pilots familiarize themselves with the controls
- Practice constant-altitude, constant-airspeed flight
- Throttle increase/decrease followed by return to a constant altitude/airspeed condition
- Straight and level 3-minute runs

Conduct Evaluation Runs

LIST OF REFERENCES

Bylander, E.G., *Electronic Displays*, McGraw-Hill, Inc., 1979.

Cooper, G.E., and Harper, R.P., NASA Technical Note TN D-5153, *The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities*, April 1969.

Crook, M.N., Hanson, J.A., and Weisz, A., WADC Technical Report 53-441, *Legibility of Type as Determined by the Combined Effect of Typographical Variables and Reflectance of Background*, March 1954.

Cushman, W. H., and Rosenburg, D. J., *Human Factors in Product Design*, Elsevier Science Publishers B. V., 1991.

Ercoline, W.R., and Gillingham, K.K., "Effects of Variations in Head-Up Display Airspeed and Altitude Representations on Basic Flight Performance," *Proceedings of the Human Factors Society*, 34th Annual Meeting, 1990.

Haworth, L.A., and Newman, R.L., NASA Technical Memorandum 103947/USAATCOM Technical Report 92-A-006, *Test Techniques for Evaluating Flight Displays*, February 1993.

Howell, W.C., and Kraft, C.L., WADC Technical Report 59-536, *Size, Blur, and Contrast as Variables Affecting the Legibility of Alpha-Numeric Symbols on Radar-Type Displays*, September 1959.

Lind, J.H., NWC Technical Memorandum 4276, *Plan for the Estimation of Pilot Workload for the AV-8B Aircraft*, November 1980.

Lind, J.H., NWC Technical Memorandum 4538, *Evaluation of Cockpit Procedures, Displays, and Controls for Stores Management in the Advanced Aircraft Armament System (AAAS)*, October 1981.

Marshall, R. W., *An Apparatus for Rapid Prototyping and Evaluation of Wide Field of View Helmet-Mounted Display Symbology*, Master's Thesis, Naval Postgraduate School, Monterey, California, December 1993.

MIL-STD-1295A, Department of Defense Military Standard MIL-STD-1295, *Human Factors Engineering Design Criteria for Helicopter Electro-Optic Display Symbology,* Revision A in preparation, January 1990.

Roufs, J.A., and Bouma, H., "Towards Linking Perception Research and Image Quality," *Proceedings of the Society for Information Display*, v. 21, pp. 247-270, 1980.

Shurtleff, D.A., and Alexander, P.J., "Legibility Criteria in Design and Selection of Data Displays for Group Viewing," *Proceedings of Human Factors Society*, November 1979.

Shurtleff, D.A., *How to Make Displays Legible*, Human Interface Design, 1980.

Snyder, H.L., "Image Quality: Measures and Visual Performance," *Flat Panel Displays and CRTs*, ch. 4, Tannas, L.E., (ed.), VNR, New York, 1985.

Spenkelink, G. P. J., and others, "An Instrument for the Measurement of the Visual Quality of Displays," *Behaviour and Information Technology*, v. 12, n. 4, pp. 249-260, 1993.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
Cameron Station
Alexandria, Virginia 22304-6145

2. Library, Code 52 2
Naval Postgraduate School
Monterey, California 93943-5002

3. Chairman, Code AA 1
Department of Aeronautics and Astronautics
Naval Postgraduate School
Monterey, California 93943-5000

4. Judith H. Lind 1
Operations Research Department
Code OR/Li
Naval Postgraduate School
Monterey, California 93943-5000

5. Dr. E. Roberts Wood 2
Department of Aeronautics and Astronautics
Code AA/Wd
Naval Postgraduate School
Monterey, California 93943-5000

6. Isaac Kaminer 1
Department of Aeronautics and Astronautics
Code AA/Ka
Naval Postgraduate School
Monterey, California 93943-5000

7. Lt. Charles Chiappetti, USN 5
Patrol Wing One
Detachment Misawa, Japan
Unit 5056
APO AP 96319-5000

8. Loran Haworth 5
NASA-Ames Research Center
Mail Stop 243-3
Moffett Field, California 94035-1000